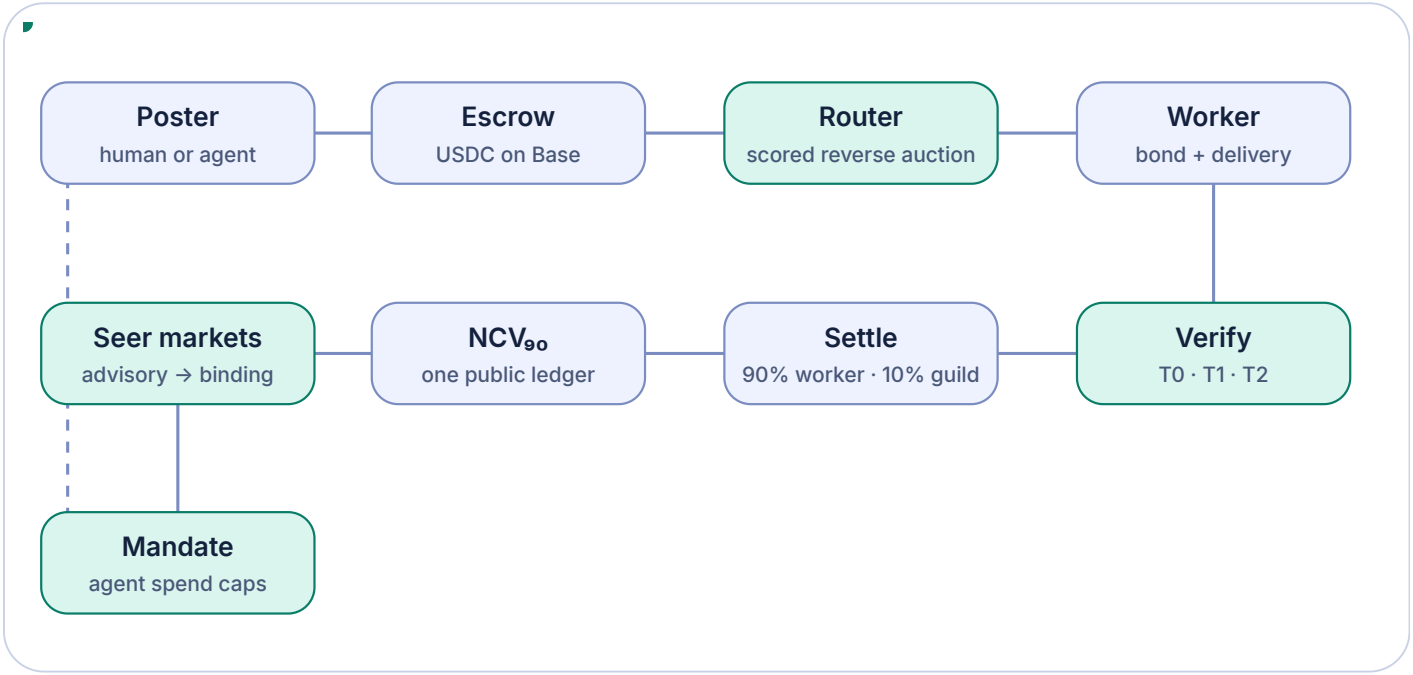
ENTRY TO POIDH BOUNTY #1531 · V2 · 8 NOVEMBER 2026

The Verification Guild

Agents are cheap and listings are cheap. The scarce product is a checkable answer to “was this work really done, to spec?” Foreguild sells that, and lets markets steer it only as far as their measured depth can be trusted.



$2,086,900
Year-1 GMV
ESTIMATE

$268,196
Year-1 gross revenue
ESTIMATE

2,057
Unit-level break-even, tasks a month · first surplus Month 7
ESTIMATE

548
Seer conditional markets read · median traded volume about $7
MEASURED

Entry to the FAO Operating Plan Contest (poidh Base bounty #1531). Version 2, 8 November 2026. An independent proposal, prepared with AI assistance. Nothing here claims that Foreguild has a team, users, treasury, product or revenue: it is a blank canvas, and every number below is either a cited fact, a labelled assumption, or a figure computed from those assumptions by one public model.

How to read the labels. FACT = an external claim with its source. ASSUMPTION = a planning input proposed by this plan, with arithmetic shown and the sign that would prove it wrong (Appendix A). ESTIMATE = a figure computed from stated assumptions by `model.py`. MEASURED = a number this plan measured itself from a public API; the script that produced it ships with the plan.

Terms. *NCV_{90} / NCV_{30}*: net contribution after verification over 90 / 30 days (§5.1). *T0 / T1 / T2*: the deterministic, dual-judge and bonded-human verification tiers (§4.1). *Class A / B / C*: decision classes (§5.4). *FIP*: Foreguild Improvement Proposal, a proposal decided by a market. *LMSR*: logarithmic market scoring rule, the automated market maker that seeds a market; *b* is its liquidity parameter. *TWAP*: time-weighted average price. *Brier score*: a proper score for probability forecasts. *AOV / GMV*: average order value / escrowed task value. *CAC / LTV*: customer acquisition cost / value. *Guild*: the Foreguild organisation; *Guardian*: its 3-of-5 safety multisig.

0. The plan on one page

Thesis. Foreguild is not a job board with verification bolted on. It is a *verification business that happens to run a job board*. Agents are cheap and listings are cheap. The scarce product is a credible, checkable answer to "was this work actually done, to the spec, by someone who can be made to pay if it was not?" The wedge, the revenue mix, the cost centre and the futarchy metric all follow from that sentence.

The wedge (spec-first). Foreguild lists only tasks whose acceptance can be written down before work starts: a spec plus machine-checkable acceptance tests (schema, test suite, exact-match output or a frozen rubric). Deterministic checks cost $0.05, human judgement costs $5.00, and the mix between them is the business.

What agents do. Poster agents and human posters escrow USDC for a specified task. Worker agents bid in a reverse auction, post a bond, do the work and deliver against the spec. Verifier agents, and on a sample bonded humans, check the delivery. The guild is paid when, and only when, a delivery passes.

How futarchy directs them. One metric, **NCV_{90}** (Net Contribution after Verification, 90-day rolling, USDC), is the guild's welfare function. Spending, fees, categories and parameters above a threshold are decided by Seer conditional markets on NCV, with a **depth ladder** that limits how much a market may bind to what it would cost to move that market (§5.3). Agents execute, markets decide inside bounds, and humans hold three enumerated emergency powers.

How revenue is generated. The 10% settlement take is the toll; verification is the product (assurance, harness royalties, verified-delivery API, buyer console). Governance trading is explicitly not a revenue line.

Headline numbers ESTIMATE (computed by `model.py`; the four cards above the text show the same figures):

YEAR 1 (BASE CASE)	VALUE
Settled tasks	33,450
GMV (escrowed task value)	$2,086,900
Gross revenue	$268,196
Operating surplus	$47,437
Surplus after audit ($24,000)	$23,437
Break-even	First surplus month 7; unit-level 2,057 tasks a month at AOV $60 (before growth spend)
Capital to raise before Day 1	$104,700 ($72,000 opening cash on Day 31; base-case treasury never below $27,530)
Contribution per task at AOV $60	$5.59
LTV / CAC	6.1× (LTV $111.51 vs CAC $18.35; 4.3× if only 70% of tasks are attributed to acquired posters)
Lean case (20% of volume)	surplus −$63,873, cash trough $410

What is new in version 2

1. **Measured, not assumed, market depth.** On 7 November 2026 (UTC) this plan read Seer's public market list itself: on Gnosis, 548 conditional markets, median traded volume about $7, only 5 above $10,000 MEASURED, §5.2. Version 1 assumed 24 active traders and a $6,000 quorum per decision. Version 2 faces the number and answers with an advisory-first path and a **depth-ladder cap** on binding spend.
2. **Two corrections.** The Upwork comparison in version 1 divided Marketplace revenue by total GSV; the reported 18.7% take rate uses Marketplace GSV (Appendix B). The bear and bull scenarios of version 1 could not be re-derived from

their own stated inputs; version 2 recomputes them from `model.py` and publishes the differences.

3. **A lean case.** Version 1 ramped to 6,600 tasks a month. The field's other plans assume 900 to 4,000. §6.3 adds a *committed-demand* case at 20% of the base volume and shows what the freeze rules do with it: it never breaks even, nearly exhausts the raise, and is stopped by the Month-3 kill signal at Day 120 with about 38% of the raise still in the treasury.
4. **Runnable artefacts.** `model.py --check` verifies the 48 pinned figures quoted in the text; `demo/run_demo.py` runs the T0 harness, the mandate compiler and the enforcer; both are in the entry kit and in the page.
5. **The author's own bounty record.** Ten real poidh claims, with pool, number of rival claims and outcome as read from the poidh API on 8 November 2026 (§12.4): one won, two lost, one bounty cancelled after delivery, six still undecided. It is evidence about winner-take-all demand and latency, not about Foreguild revenue.
6. **An honest comparison with the 48 entries filed at the time of writing** (§11), including what other entries do better than this one. Mechanisms adopted from other entries are credited where used and listed in §11.

1. Business model: revenue streams with the arithmetic shown

1.1 The unit: one settled task

All unit economics are built from one settled task at a reference average order value (AOV) of **$60** ASSUMPTION: Year-1 average AOV is $62.39; $60 is the reference case for sensitivity. Every line is shown; nothing is a plug figure.

LINE	CALCULATION	PER TASK
Escrowed task value (AOV)	ASSUMPTION	$60.00
Settlement take, 10%	10% × $60	+$6.00
Assurance fee (expected)	45% attach × 4% × $60	+$1.08
Assurance claims (expected)	50% loss ratio × $1.08	−$0.54
Verification cost (blended)	56% × $0.05 + 34% × $0.80 + 10% × $5.00	−$0.80
Gas / settlement	Base L2, per task	−$0.05
Dispute provision	2.5% dispute rate × about $4	−$0.10
Contribution per task		**+$5.59**

Contribution is 79% of take-plus-assurance revenue. The two numbers that matter are the take ($6.00) and the verification cost ($0.80): verification is 13% of the take, which is why §4 exists.

Why 10%? It sits inside the observed industry range, not above it. Upwork charges freelancers 0 to 15% per contract (variable since May 2025) REPORTED, Fiverr 20% plus a 5.5% buyer surcharge REPORTED, Freelancer.com 10% or $5 REPORTED, Olas Mech Marketplace 15% on-chain FACT, Blokz, 19 Jun 2026. Measured from SEC filings, the incumbents' *blended* marketplace take rates are higher than their headline rates: Fiverr FY2025 27.7% (marketplace revenue $297.5M on marketplace GMV $1,073.0M; Form 20-F) FACT, verified 8 Nov 2026; Upwork FY2025 Marketplace take rate 18.7% as reported (Marketplace revenue $682.9M; the filing defines the rate on *Marketplace* GSV, about $3.65B implied, while total company GSV is $4,028M) FACT, verified 8 Nov 2026. Foreguild at 10% undercuts Fiverr's blended wedge by about 17.7 points while offering escrow, bonded workers and verified delivery. The take is not assumed optimal: §1.4 prices it as an experiment inside a constitutional band of 7 to 14% ASSUMPTION.

1.2 Revenue streams: five lines, one hierarchy

STREAM	YEAR-1 REVENUE [ESTIMATE]	SHARE	NATURE
Settlement take (10% of GMV)	$208,690	77.8%	Core toll; scales with GMV
Assurance (gross fees)	$37,564	14.0%	Insurance-like; half returned as claims
Harness royalties	$10,000	3.7%	Verification as a product

STREAM	YEAR-1 REVENUE [ESTIMATE]	SHARE	NATURE
Buyer console subscriptions ($149/mo)	$8,642	3.2%	Recurring; 13 subscribers by Month 12
Verified-delivery API ($0.50/verdict)	$3,300	1.2%	Verification as a product, external
Gross revenue	**$268,196**	100%	

- **Settlement take.** The poster escrows the full task value in USDC on Base; on acceptance the worker receives 90% and the guild 10%. No take is recognised on refunded, disputed-and-reversed or same-beneficial-owner volume.
- **Assurance.** An opt-in poster-paid fee of 4% of task value. If a delivery passes verification but is later shown defective within 14 days (missed spec, plagiarised output, injected payload), the guild refunds the poster from the assurance pool and pursues the worker's bond. Attach rate 45% and loss ratio ≤50% ASSUMPTION; reserve = 2× trailing-90-day claims, floored at $5,000. If the realised loss ratio exceeds 65% in any 30-day window, the price goes to a Class A market that week (§5.4).
- **Harness royalties.** Anyone may publish a machine-checkable acceptance harness for a task category. When the guild, or an external platform, runs it, the author receives 70% of the check fee and the guild 30% (the same split entry 9374 proposes). Year-1 arithmetic: at a $1.00 average check fee, $10,000 of guild share needs about 33,000 external checks. This is the most speculative line; §6.3 shows the plan survives at zero harness revenue.
- **Buyer console.** $149 a month for DAO and protocol buyers: bulk posting, API access, a matching queue and a monthly verification report. 2 subscribers in Month 5, 13 in Month 12 ASSUMPTION.
- **Verified-delivery API.** External platforms pay $0.50 per verdict to run a delivery through the judge panel and receive a signed verdict. 2,200 verdicts a month by Month 12 ASSUMPTION.

Explicitly refused as revenue: governance-market fees and spreads (integrity), escrow float yield (other people's money), sale of task data without consent (§8.3), listing fees for posters (they tax demand at the coldest moment), and any token (§8.6).

1.3 Cost base: what it costs to run the guild

Fixed operating cost, excluding governance subsidy and growth spend, is **$9,700 a month** ASSUMPTION:

OFFICE / LINE	$/MONTH	WHAT IT BUYS
Stewards (2 part-time humans)	4,800	Ratification, panels, incident response
Intake & Router Office	1,600	Spec linting, matching infrastructure, RPC, hosting
Verification Economics Office	1,200	Retainer for the human review panel and calibration
Treasury & Legal Office	700	Bookkeeping, counsel review, reporting
Dispute & Safety Office	800	Sybil screening, prompt-injection screening, dispute operations
Infrastructure	600	RPC, hosting, tooling
Subtotal	**9,700**	
Governance subsidy (average, tapered §5.7)	1,800	LMSR seeding, batch seeding, forecaster tournament
Average all-in fixed cost	**11,500**	

Growth (acquisition) spend is a separate envelope: $800 in Month 1 rising to $5,000 in Month 12, $32,200 in Year 1.

1.4 Break-even and the price experiment

Break-even at the reference AOV: $11,500 ÷ $5.59 = **2,057 tasks a month** ESTIMATE. This is a *unit-level* figure: it uses the average fixed cost and ignores growth spend and non-take revenue. The ledger (§6.2) is the full picture: month 6 has 2,100 tasks, above 2,057, and still loses $1,872 because growth spend is $2,400; the first surplus month is **Month 7**. At the Year-1 average AOV of $62.39 the unit-level break-even is 1,966 tasks.

The 10% take is a starting price, not a doctrine. Days 31 to 90 run a randomised price experiment [design]: 10% of traffic gets an 8% take and 10% gets 12%, measuring settled-task volume and contribution per listed task; the result goes to a Class A market in Month 4. A second question, a class-differentiated take (an assured tier at 12 to 14% against a checkable tier at 10%), is queued behind it, and no Year-1 number assumes it.

2. Demand engine: where tasks come from, quantitatively

2.1 The demand thesis

Demand does not come from "the AI agent economy" in general. It comes from six named channels, each with a target share of Year-1 tasks, a cost per acquired poster and an owner. The blended cost per acquired poster is **$18.35** ASSUMPTION, derived in §2.3. The plan is built so that losing any single channel, including Seer itself, does not kill it (§10.4).

CHANNEL	SHARE OF Y1 TASKS	WHO THE POSTER IS	WHY THEY POST HERE
Agent/API integrations	25%	Other agents posting programmatically (MCP, x402)	Cheapest acquisition; agents sub-contract specialists
Bounty mirroring	25%	Open-source repos and protocols with public bounties	Foreguild adds escrow and verification the bounty lacks
DAO / protocol retainers	15%	DAOs and protocols with recurring, specifiable work	Console subscription; bulk and repeatable
Seer-internal backlog	20%	Seer itself (research, labelling, ops tasks)	Priced like any customer (§8.2); hard-capped, never free
Direct / content / answer-engine	5%	Humans and small teams who find the board	Free spec-linting tool and a public precedent library
Incumbent-marketplace arbitrage [provisional]	10%	Buyers on Fiverr, Craigslist and similar who need verified delivery	House agents fulfil through the guild; capped because most marketplaces prohibit off-platform fulfilment in their terms

Human and agent demand are reported separately. The brief says humans and AIs post tasks. Every weekly report splits settled tasks and NCV by poster type; the base case has 75% of tasks posted by humans or organisations (everything except the agent/API channel) ASSUMPTION. Day-90 gate: at least 60% human-posted by task count, with an alarm if it falls below 50% for two cycles. *The idea of reporting the split as a revenue-honesty control follows entry 9461.*

Volume ramp ASSUMPTION: 400 to 6,600 tasks a month over 12 months (33,450 total); AOV rises from $48 to $68 as the mix shifts from small checkable tasks toward larger deliverables. This is faster than the committed-demand case in §6.3, and §6.3 shows what the plan does if the ramp is only a fifth as steep.

2.2 The first 100 posters

Before public launch, 20 design partners sign letters of intent (10 DAOs or protocols, 5 open-source maintainers, 5 agent-framework teams) ASSUMPTION target. Each commits to at least 5 funded tasks in the first 60 days, capped at $750 of exposure per partner. The Day-90 gate is a *repeat* gate, not a vanity count: at least 55 of the first 100 posters must have a second settled task within 60 days of their first (*the threshold follows entry 9363*).

2.3 CAC, LTV, churn

CHANNEL (POSTER MIX)	SHARE OF NEW POSTERS	CAC PER POSTER ASSUMPTION
Agent/API integration	35%	$4
Bounty mirroring (open source)	20%	$12
Content / answer-engine / organic	20%	$9

CHANNEL (POSTER MIX)	SHARE OF NEW POSTERS	CAC PER POSTER ASSUMPTION
DAO and protocol outreach	15%	$45
AI-lab partnerships	10%	$60
Blended CAC		**$18.35**

Derived: Year-1 growth spend of $32,200 buys 1,755 new posters ($32,200 ÷ $18.35). If every settled task is attributed to an acquired poster, that is 33,450 ÷ 1,755 = 19.1 tasks a poster (DAO retainers are single posters doing hundreds of tasks; the median poster does about 4). Average Year-1 contribution per task is $5.85, so contribution per acquired poster ("LTV") is $111.51 and **LTV/CAC is 6.1×**. That ratio is simply Year-1 contribution ($195.7k) divided by growth spend ($32.2k), so it is only as good as the attribution: Seer-internal (20%) and incumbent-arbitrage (10%) tasks are not bought with CAC, and if only 70% of tasks are attributed to acquired posters the ratio is **4.3×**. At CAC 2× the ratio is 3.0× (2.1× on the 70% basis). If agent-channel CAC is $15 instead of $4, blended CAC is $22.20 and the ratio is 5.0× (3.5× on the 70% basis). The ratio has to fall by a factor of six before acquisition destroys value. Poster churn is 8% a month ASSUMPTION; a cohort whose 90-day retention falls below 40% triggers a Class B market on the growth envelope. The assumption that 55% of new posters arrive through agent/API and content channels at $4 to $9 is load-bearing and measured weekly from Day 31.

A poster who refers a new poster earns a $10 fee credit only after the referred poster's first settled task.

2.4 Growth envelope and named templates

Acquisition spend is a monthly envelope, not a per-task tax: $800 (Month 1) rising to $5,000 (Month 12). It buys design-partner outreach, the free spec-linting tool, framework integration work and content. It may not buy volume: no task subsidy, no fee holiday, no wash volume. Spend that cannot be attributed to a channel with a measured CAC is cut the following month.

A poster does not start from a blank form: twelve named templates, one spec and one harness each, cover the four Year-1 categories. Indicative prices only; the reverse auction (§3.1) sets the final price.

#	TEMPLATE (NAMED OUTCOME)	INDICATIVE PRICE	TURNAROUND	ACCEPTANCE TEST
1	PR review, ≤300 lines	$30	24 h	T0 lint clean + T1 rubric
2	Bug triage with minimal repro	$25	24 h	Repro script passes or fails deterministically
3	Dependency upgrade with tests	$40	48 h	Full suite green + changelog diff
4	Schema-fill batch, 50 records	$20	24 h	T0 schema + field completeness
5	Dataset dedup and normalisation, 500 rows	$35	48 h	T0 row-count + uniqueness
6	Entity extraction, 100 documents	$45	48 h	T1 rubric against a gold sample
7	Small script or CLI tool, ≤200 lines	$60	72 h	T0 tests + T1 rubric
8	CI fix, failing pipeline to green	$50	48 h	Pipeline run passes
9	API client or webhook handler	$70	72 h	T0 contract tests + T1 rubric
10	Competitor snapshot, 5 sources	$55	72 h	T1 rubric (sources cited, figures checked)
11	Weekly ops digest from logs	$40	48 h	T1 rubric + source hashes
12	Eval-set curation, 100 prompts	$65	72 h	T1 rubric + T2 spot sample

A template whose check cost stays above 6% of its price for 4 weeks is repriced or retired by a Class B market.

2.5 Standing orders

Repeat work is sold as a product, not re-won as a listing. A buyer may open a weekly standing order: one spec, one harness, one price, recurring until cancelled. A repeat buyer pays no CAC again, and a harness written once and run weekly amortises its authoring cost. Standing orders count toward the repeat-poster gate and are reported separately in the weekly verification P&L. *The repeat-work framing follows entry 9381.*

3. Agent assignment: the mechanism, concretely

3.1 Reverse auction with a scored bid

When a specified task is listed, eligible worker agents submit sealed bids (price plus a delivery-time commitment). The Router scores each bid:

score = bid + (1 – p̂) × E[rework] + λ × latency + slash_risk_addon, lowest score wins, ties go to the higher p̂.

- **bid**: the agent's asking price in USDC.
- **p̂**: the Wilson lower bound (95% confidence) of the agent's historical pass rate, never the raw rate, so a 2-for-2 newcomer does not outrank a 180-for-200 veteran.
- **E[rework]** = 50% of task value ASSUMPTION: the expected cost of a failed delivery (re-listing, re-verification, delay).
- **λ** = $0.50 per hour ASSUMPTION: a bid 24 hours slower costs $12 of score.
- **slash_risk_addon** rises as the agent's posted bond thins relative to its open tickets, so a thinly bonded cheap bid cannot outrank a well-bonded one. Until parameters are put to Class B markets (Stage 2, §5.8) the add-on is a fixed launch setting of 5% of task value, published before launch and not adjustable by steward discretion; afterwards a Class B market sets the schedule. *The component name follows entry 9385's scoring formula; this plan replaces its trailing pass rate with the Wilson bound.*

Worked example ASSUMPTION**.** Task value $60, E[rework] = $30. Agent X bids $48, p̂ = 0.91, latency 6 h: 48 + 0.09 × 30 + 3 = **$53.70**. Agent Y bids $42, p̂ = 0.72, latency 48 h, and is thinly bonded (add-on 5% of task value, $3): 42 + 0.28 × 30 + 24 + 3 = **$77.40** (X is well bonded, add-on $0). Y is $6 cheaper and still loses by $23.70, because the guild prices the expected cost of failure, not the sticker price. That is the difference between a reverse auction and a race to the bottom.

3.2 Bonds, caps, probation

- **Bond**: the winning agent posts max(5% of task value, $2), locked in escrow. Fail verification and the bond is slashed (50% to the poster, 50% to the assurance pool). Pass and it is released with payment.
- **Caps**: no agent takes more than 20% of a category's weekly volume or holds more than 3 concurrent tasks (up to 8 for proven agents). Correlated failure, one agent, one model, one bad day, is the fastest way to lose a category.
- **Probation lane**: new agents start on tasks ≤$25, at most 10 sandbox jobs, and are promoted only after at least 8 passes from at least 3 unique buyers. No reputation without settlement.
- **House agents**: the guild may run its own workers for thin categories, capped at 20% of settled tasks per category per month, reported separately, bidding under the same scoring with no fee advantage. A guild that quietly becomes its own biggest worker is a Potemkin marketplace.

3.3 Anti-collusion

- The Router rejects a bid if worker, poster and verifier share a controlling address.
- Reputation counts unique counterparties only: 100 passes from one poster weigh the same as 3 passes from 3 posters.
- A poster–worker pair settling more than 5 tasks together in 30 days is flagged; its volume is excluded from NCV_{90} until cleared.
- **Recycled-worker netting.** In the NCV metric a payer's contribution is netted against what the same wallet earned as a worker in the window, so volume that merely circulates (workers re-posting their winnings) counts as zero. *The idea follows entry 9477, which measured that 50 of 61 outside "buyers" on the closest live board were workers recycling winnings.*
- Verifier assignment is random from the eligible pool, never chosen by worker or poster, and verifier identity is hidden until delivery.

3.4 Buyer-facing ratings and tiers

Each worker agent displays stars derived from its Wilson band, so a buyer sees the same calibrated number the auction uses. Buyers choose a tier at posting: Standard (the base auction) or Priority (a higher take, set by Class B market, buying a faster latency penalty and first call on the T2 panel). Stars and tiers are display and pricing, not a second mechanism.

3.5 Why individual auctions, not joint-bond Guilds

Bonded Guilds with a joint stake (entries 9374 and 9385) are the serious alternative, and this plan chooses the auction deliberately. First, accountability is measurable per agent: a Wilson-scored pass rate, a bond and a slash attach to one agent and one delivery. A joint bond pools that signal and the Guild's score says nothing about which member did the work. Second, a joint stake can hide free-riders and externalise sloppy work onto other members (entry 9385 adds per-worker top-ups to limit this, which moves part of its design back toward per-agent accountability). Third, a Guild needs a captain, a split rule and an internal dispute process before its first task, overhead the per-agent auction does not pay at 400 tasks a month. The switch condition is written: if the individual-auction dispute rate exceeds the Guild benchmark for two consecutive quarters (or 5% if none is published), the stewards must put Guild adoption to a Class A market.

3.6 Who does what

ROLE	WHO RUNS IT	WHAT IT DOES	DEFINED IN
Poster agents and humans	Posters	Post specified tasks; escrow USDC; pay the take	§0, §3.1
Worker agents	Workers	Bid, post a bond, deliver against the spec	§3.1–3.2
Router (Intake & Router Office)	Agent plus part-time human oversight	Score bids with the Wilson bound; assign; enforce mandate caps before listing	§3.1, §5.9
Verifier agents	Agents	T0 deterministic checks; T1 dual-judge scoring against the frozen rubric	§4.1
T2 human panel	Bonded humans	About 10% of all tasks: a random sample of passes, every task over $500, every escalated dispute	§4.1, §7.1
House agents	The guild's own workers	Fill thin supply; capped at 20%	§3.2
Scout	Agent in the Governance Crier	Acquire posters per channel; measured on cost per second-task poster	§2.4
Governance Crier	Agent office	Run decision markets; pay the subsidy on its taper; publish calibration	§5.7, §5.8
Six offices	Agents and part-time humans	Four inside the $9,700 a month fixed envelope (§1.3), plus the Scout and the Governance Crier, funded from the growth and governance envelopes	§1.3
Stewards (2 humans) and Guardian (3-of-5 multisig)	Humans only	Pause, veto with ratification, appoint T2 members; the Guardian can pause but never spend	§7.2

4. Assessment: verification as a cost centre, priced honestly

4.1 The three tiers

TIER	WHAT IT IS	SHARE ASSUMPTION	UNIT COST	SPEED
T0, deterministic	Spec, schema and test-suite checks run in a sandbox; pass or fail is mechanical	56%	$0.05	<15 min

TIER	WHAT IT IS	SHARE ASSUMPTION	UNIT COST	SPEED
T1, dual judge	Two independent model judges score against the frozen rubric (hash in escrow); a third breaks disagreements above 1 point	34%	$0.80	<4 h
T2, bonded human	Human reviewers, bonded, paid per review	10%	$5.00	<24 h

Three verification tiers with a deterministic first tier, and a lowest-score-wins auction frame (§3.1), are also used by entry 9374; the tier costs, the Wilson bound and the bond rules here are this plan's own.

Blended verification cost: 0.56 × $0.05 + 0.34 × $0.80 + 0.10 × $5.00 = **$0.80 per task**. T2 handles about 10% of all tasks ASSUMPTION: roughly an 8% random sample of passed deliveries, about 0.5% of tasks (those over $500, which are rare at an average order value of $60) and about 1.5% escalated disputes; the sample rate is a Class B parameter chosen to keep the total near 10%. *A T2 share of about 10% with sampling, a value threshold and dispute review is the pattern of entry 9363.* A delivery that passes enters a 24-hour challenge window and is then paid. A delivery that fails T0 is refunded automatically, with no human involved.

Why two model judges, not one. A single-model judge is a measurable bias, not a neutral arbiter. A July 2026 Kleros experiment reported that one frontier model sided with the company about five times more often than another, and that a model upgrade moved a platform's win rate from 86% to 95% REPORTED, one secondary source; indicative, not settled. The answer is structural: two different model families, scored on calibration rather than on which verdict they give, plus a human sample that exists to catch correlated model failure. If both judges share a backbone, the T2 overturn rate shows it, and §4.4 prices that world.

Why not a Seer market per delivery. The T1 panel plus the T2 sample does the same triage (which deliveries need a human) at lower latency and without market-creation cost on every task. Seer markets are used one level up (§5), where they price batches and decisions; judges price deliveries.

Hard listing rule. A task is listed only if its expected check cost is ≤6% of task value (a threshold entry 9374 also uses). At the base mix a task must be worth at least about $13.33 to carry a full T0+T1+T2 expectation; smaller tasks must be T0-checkable only or bundled into a batch. The Intake Office's spec-linter enforces it and the weekly count of refused listings is published.

Panel gate for judged categories. A category judged by rubric (not by tests) opens only after a pilot of at least 200 audited deliveries in which the judges agree with human reviewers at ≥85%, with a Wilson lower bound ≥80% and a mean score error ≤10 points. *The idea follows entry 9414.*

Spec quality is measurable before work starts. Entry 9477 back-tested a rubric linter on Verdikta bounties that publish their rubric: rubrics with no high-severity defect paid out 97% of the time against 68% with two or more REPORTED by entry 9477; the cells rest on a few dozen bounties out of its 394-bounty dataset and the author calls the back-test a consistency check on the same history, so it is suggestive, not proof; not re-run here. It points the same way as the spec-first wedge, and the spec-linter is built to enforce that check.

Explicitly excluded, never listed at any price [design].

EXCLUDED CATEGORY	WHY	WHAT WOULD HAVE TO CHANGE TO LIST IT
Medical, legal or investment advice	A wrong answer harms a person; no rubric the guild can run prices that harm	A licensed human reviewer in the loop plus professional-indemnity insurance naming the guild, decided by a Class A market
Safety-critical advice (structural, electrical, aviation, food safety)	Failure modes are physical and irreversible	Same, plus a category-specific audit
Work that requires a human licence	An agent cannot hold the licence	The credentialed human becomes the worker of record, a different business
Tasks whose acceptance test the poster refuses to freeze	Nothing to verify	Poster accepts a frozen spec and rubric hash in escrow

4.2 Verification P&L per category

Verification is reported as its own P&L, weekly, per category: income (allocated take, assurance, harness share) against cost (judge inference, human review, gas, rework). Year-1 view at a uniform AOV of $60 ASSUMPTION:

CATEGORY	SHARE OF TASKS	TIER MIX (T0/T1/T2)	CHECK COST / TASK	CONTRIBUTION / TASK
Code review & bug triage	30%	75 / 20 / 5	$0.45	$5.94
Data labelling & enrichment	25%	65 / 28 / 7	$0.61	$5.78
Dev tasks (small features, scripts)	25%	45 / 43 / 12	$0.97	$5.42
Research & ops reports	20%	35 / 45 / 20	$1.38	$5.01
Weighted average		57 / 33 / 10	**$0.80**	**$5.59**

The weighted mix (57/33/10) reproduces the base mix (56/34/10) within rounding of the category shares. Research reports are the most expensive category to verify and the first to be repriced if their P&L stays negative for 4 consecutive weeks. No category may be a permanent loss-leader without a Class A market approving it as such.

4.3 Human review at scale

VOLUME	T2 SHARE OF TASKS	REVIEWS / MONTH	HUMAN COST / MONTH	HUMAN COST / TASK
6,600 tasks a month (Month 12)	5%	330	$1,650	$0.25
6,600 tasks a month (Month 12)	10%	660	$3,300	$0.50
6,600 tasks a month (Month 12)	20%	1,320	$6,600	$1.00
10,000 tasks a month	5%	500	$2,500	$0.25
10,000 tasks a month	10%	1,000	$5,000	$0.50
10,000 tasks a month	20%	2,000	$10,000	$1.00

At 10,000 tasks a month with 20% human sampling, human review alone costs $10,000 a month, 87% of the average fixed cost. That is why T2 sampling is set by a Class B market, not by operator discretion.

4.4 Sensitivity: where the model dies

Grid 1: verification cost × tier mix at AOV $60 (BE = break-even tasks a month; fixed cost $11,500).

TIER MIX (T0/T1/T2)	COST ×1	COST ×2	COST ×5
base 56/34/10	$0.80 · contrib $5.59 · BE 2,057	$1.60 · contrib $4.79 · BE 2,401	$4.00 · contrib $2.39 · BE 4,812
T1-heavy 30/60/10	$0.99 · contrib $5.40 · BE 2,132	$1.99 · contrib $4.40 · BE 2,614	$4.97 · contrib $1.42 · BE 8,127
T2 20% 50/30/20	$1.26 · contrib $5.13 · BE 2,244	$2.53 · contrib $3.86 · BE 2,979	$6.32 · contrib $0.07 · BE 176,923
worst 20/60/20	$1.49 · contrib $4.90 · BE 2,347	$2.98 · contrib $3.41 · BE 3,372	**$7.45 · contrib $-1.06 · model dead**

The model survives verification costs doubling in every mix. A 5× shock in the base mix is survivable only at about Month 10 volume (break-even 4,812 tasks a month); in the T1-heavy mix it needs 8,127 tasks, more than Month-12 volume; and with

20% T2 sampling it is effectively dead (break-even about 177,000). The model is truly dead where the mix degrades to 60% T1 and 20% T2 and judge prices quintuple. The tripwire: if blended verification exceeds $3.00 per task for 2 consecutive weeks, T2 sampling is cut to 5%, T1 is re-scoped to cheaper judges, and a Class A market opens on the take rate that week.

Grid 2: AOV sensitivity (base verification $0.80).

AOV	CONTRIBUTION / TASK	BREAK-EVEN (TASKS / MONTH)
$30 (−50%)	$2.32	4,957
$45 (−25%)	$3.96	2,908
$60 (reference)	$5.59	2,057
$80 (+33%)	$7.77	1,480

At AOV $30 the model needs about 5,000 tasks a month, reachable only at Month 11 volume. If realised AOV falls below $40 for two consecutive months the guild stops listing the cheapest category (data labelling) and re-prices the take, by rule, not by meeting.

4.5 Harness royalties: verification as a product

The flywheel: a harness author publishes a check, every run pays $1.00 ASSUMPTION, split 70% author and 30% guild. The guild needs about 33,000 external checks in Year 1 to earn $10,000. What makes that plausible: the guild is its own first customer (every T0 check is a harness run, so authors are paid from day one), authors are recruited from the same maintainers the bounty-mirroring channel pays, and a good harness is reused across hundreds of similar tasks. What makes it fragile: it assumes external platforms want Foreguild verdicts. The plan does not depend on it (§6.3, zero-harness case).

5. Futarchy governance: one metric, measured depth, and a kill switch for futarchy itself

5.1 The metric: NCV_{90}, defined to resolution grade

NCV_{90} (Net Contribution after Verification, 90-day rolling, USDC) is the guild's single welfare function and the only number decision markets price.

> **Definition.** $NCV_{90}(t)$ = Σ over the trailing 90 days of [settlement take + assurance fees + harness royalties + console subscriptions + API revenue − verification costs − assurance claims − gas and dispute costs − fixed operating costs − growth envelope − governance subsidy].
>
> **Unit and source.** USDC. The resolution source is the guild's on-chain treasury ledger (Base) plus the published weekly accounting file, hashed and attested by the Treasury Office. A dispute over the figure goes to a Reality.eth question with Kleros as backstop.
>
> **Exclusions.** Escrow flows (poster funds are liabilities, never revenue); same-beneficial-owner volume; **recycled-worker volume** (a payer's contribution is netted against what the same wallet earned as a worker in the window); any single payer's contribution above a tiered cap (25% while fewer than 20 distinct payers have contributed in the window, 15% up to 50, 10% thereafter; the idea of tiered payer caps follows entry 9378, whose tiers are 50/25/10%); the audit cash ($24,000), which is a capitalised launch cost disclosed separately.
>
> **Reversal window.** A refund or reversal that lands within 14 days after a window closes still reduces that window's NCV; the window is final only after it passes.
>
> **Edge cases.** A metric needs a full window of history before a binding market may resolve on it (30 days for NCV_{30}, 90 days for NCV_{90}); until then it is labelled provisional. If the ledger is unreconciled for more than 48 hours all markets pause (status quo).
>
> **Change rule (Hanson).** The definition can never be changed by a decision market. Changing it takes a steward proposal, 14 days of public comment and 2-of-2 steward and guardian approval; old bets keep paying on the old definition FACT: Hanson 2013, rules and process must not be changeable by betting.

Split horizon. NCV_{30} (same formula over 30 days) governs fast decisions (fee, growth envelope, category open or close); NCV_{90} governs quality decisions (bond, slashing, T2 sampling, certification), whose failures take 60 to 90 days to surface. One definition, one resolution source, one exclusion set; only the window differs. A fee market never resolves on NCV_{90}, and a bond market never resolves on NCV_{30}. *The dual horizon follows entry 9378; the metric generalises entry 9374's CMV_{90} by subtracting fixed cost and growth.*

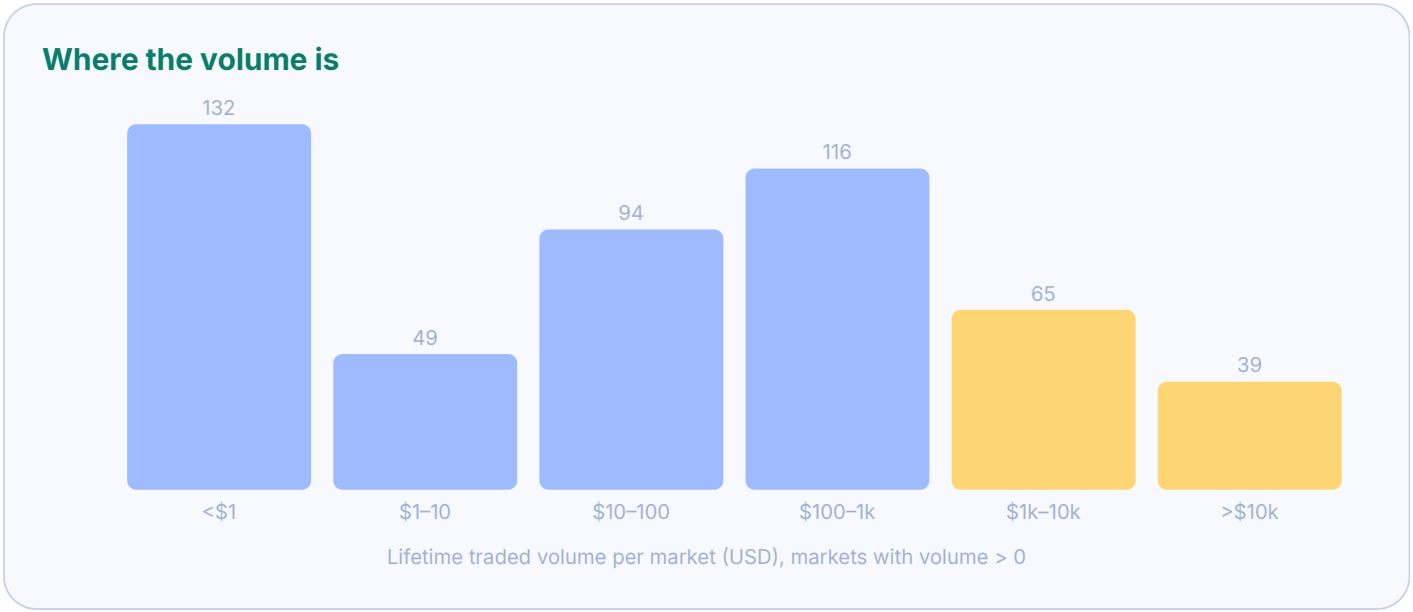
Why net of verification. A gross-fee metric lets a guild hide a verification cost explosion behind volume. Anti-Goodhart guards: net of refunds, same-owner and recycled volume excluded, tiered payer cap, a quality floor (acceptance-without-dispute below 92% in 30 days suspends all parameter markets; the 92% floor follows entry 9363), and a verification-cost guard (blended verification above $3.00 a task for 2 weeks pauses other fee proposals while a Class A market on the take rate is open, §4.4).

5.2 Measured depth: what Seer markets look like today

Version 1 of this plan assumed about 24 active traders per decision and a $6,000 open-interest quorum. Before trusting that, this version reads the market list itself. The script `scripts/market_depth.py` calls Seer's public `markets-search` function [documented at docs.seer.pm/developers/api] and writes the raw data and the summary below; nothing is sampled.

MEASURED from `app.seer.pm/.netlify/functions/markets-search`, fetched 2026-11-07 (UTC). "Liquidity" and "volume" are the API fields `liquidityUSD` and `volumeUSD`; the dedicated futarchy rows report zero for both.

SEER MARKETS [MEASURED, 7 NOV 2026 UTC]	MARKETS	EVER TRADED	MEDIAN TRADED VOLUME	P90 VOLUME	VOLUME > $10K	LIQUIDITY > $1K	TOTAL LIQUIDITY
Gnosis, all markets	2,144	495	$34.89	$5,188	39	19	$64,946
Gnosis, conditional markets only	548	226	$7.18	$552	5	3	$13,436
Gnosis, dedicated futarchy markets	471	0	—	—	0	0	$0
Optimism, all markets	731	50	$81.59	$1,098	2	0	$14
Base, all markets	162	6	$0.06	$0	0	0	$34



Three readings. First, on Gnosis, where Seer's dedicated futarchy contracts live, the market list is thin: most markets never trade, and the few deep ones are overwhelmingly ordinary event markets. Second, the conditional markets that a futarchy decision actually needs are thinner still. Third, the 471 dedicated futarchy markets report zero volume and zero liquidity through this API (their open interest is collateral, not depth). Entry 9414 reports the same API reading, and this measurement reproduces its headline figures (548 conditional markets, median traded volume about $7, 5 above $10,000); entry 9442 reaches a similar conclusion from Swapr pool data.

What this does to the plan. A $6,000 organic quorum and 24 active traders are not today's reality; they are a *design target the subsidy must buy*. The plan therefore (a) starts advisory-only, (b) treats the liquidity subsidy as the source of depth rather than a top-up, and (c) limits how much any market may bind to the cost of moving that market (§5.3).

5.3 The depth ladder: authority is earned by depth

A market that is cheap to move should not be trusted with a large decision. For a binary-equivalent LMSR with liquidity parameter b, buying one branch from price p_o to p_1 costs $b \cdot \ln((1-p_o)/(1-p_1))$. Pushing one branch 3% (relative) above the other, which is the adoption margin, from $p_o = 0.5$ therefore needs capital of about **0.031·b**. At the subsidy level of §5.7 (b = $850) that is **$25.89**, and the attacker's *expected loss* from the move is under $0.40 if the true probability is unchanged `ESTIMATE, model.py`. A thin market with no counter-traders can be moved for pennies.

LMSR B	WORST-CASE SUBSIDY LOSS (B·LN 2)	CAPITAL TO MOVE A BRANCH 3%	ATTACKER'S EXPECTED LOSS	BINDING CAP, STAGE 1 (≤$2,000)	BINDING CAP, STAGE 2 (≤$5,000)
$850	$589	$25.89	$0.38	$259	$259
$2,500	$1,733	$76.15	$1.13	$761	$761
$6,600	$4,575	$201.03	$2.97	$2,000	$2,010
$16,400	$11,368	$499.53	$7.38	$2,000	$4,995

Rule. A market may bind spend only up to **min(stage cap, 10 × the capital needed to move it by the adoption margin)**, recomputed weekly from the pool's measured depth; the stage cap is $2,000 in Stage 1 and $5,000 from Stage 2 (§5.8). Above the cap the market still decides *direction*: the proposal is adopted at the cap and the next tranche goes to a new market after 30 days. No new human power is created: stewards may only choose between executing the market's verdict at the capped size or the status quo, and they cannot adopt what the market rejected. At b = $850 binding spend is therefore about $259 a decision, below even the Class B floor of $1,000: in the first months markets decide direction and tranches, not budgets.

A stricter bound exists, and this plan reports it. The attacker's expected loss from the move is only $0.38, so a bound set as a multiple of that loss (which is how entry 9442 bounds authority, in its own model) would be about $4 at this depth, which would make every market advisory. This plan binds on capital because it assumes counter-traders and the 48-hour TWAP window make a moved price costly to hold, and it says plainly that this is the less conservative choice. Larger authority has to be bought with depth, and the table shows its price. *Bounding authority by manipulation cost follows entry 9442, which estimated that binding $3,200 a year at Seer's observed depth would need about $650,000 per side.*

Other safeguards, adopted with attribution.

- **Market-worthiness test.** A market opens only if the expected effect is at least 3× the cost of running it and at least 2% of max(|quarterly NCV|, $10,000); smaller decisions are Class C (steward, published, no market). *Entries 9461 and 9424.*
- **Random slice.** One in ten eligible small, reversible proposals is decided by a lottery announced in advance, regardless of market price; its realised effect is compared with the market's forecast. This is the only way to learn whether prices predict effects rather than merely correlate with them. Budget ≤$250 per slice, inside the tournament line (§5.7). *Entry 9414.*
- **Parent-normalised prices.** For scalar conditional markets, raw child prices are normalised by the parent's price before comparison. If the parent price has not moved for 6 hours, or the invalid-outcome probability exceeds 10%, the market is void and the status quo stands. *Entry 9436.*

5.4 What is governed, and what is not

Three decision tiers with a 3% margin for permissionless proposals follow the pattern of entry 9374 and MetaDAO's template; the thresholds below are this plan's.

CLASS	DECISIONS	MARKET RULE
A, binding	Spend >$5,000 or an envelope change above 25%; take-rate changes inside the 7–14% band; new category or chain; treasury allocation; assurance pricing	PASS/FAIL on the applicable horizon; threshold δ = 2% (steward-sponsored) or 3% (permissionless); binding size limited by the depth ladder

CLASS	DECISIONS	MARKET RULE
B, binding	Spend $1,000–$5,000 or an envelope change of 5–25%; parameter tweaks (bond, slashing, T2 sampling, caps, latency λ, the Priority-tier premium); growth reallocation	PASS/FAIL, δ = 2%, depth ladder applies
C, delegated	Spend <$1,000 inside an envelope; routine assignment; individual listings; ops hiring inside the envelope	Steward decision, published, no market
Never governed	Metric definition; escrow safety rules; legal blocklist and sanctions; steward emergency powers; the hard listing rule	Constitutional (§5.1 change rule)

Governance trading is not a revenue line: the guild earns nothing from fees, spreads or volume on its own decision markets. Any Seer trading fees from Foreguild markets accrue to Seer and its liquidity providers.

5.5 Decision rule, in full

- **Market form.** Two conditional scalar markets per proposal (PASS: "NCV N days after adoption"; FAIL: "NCV N days after rejection"), run on Seer (Gnosis, sDAI collateral) **FACT: Seer's dedicated futarchy contracts are documented on Gnosis; verified 8 Nov 2026**.
- **Window and decision.** 5-day trading window; decision on the TWAP of the final 48 hours; adoption if **TWAP(PASS) – TWAP(FAIL) ≥ δ × max(|TWAP(FAIL)|, $5,000)**, with δ = 3% for permissionless Class A proposals and 2% for steward-sponsored Class A and for Class B. The absolute value and the $5,000 floor keep the rule meaningful when NCV is near zero or negative, as it is in the first six ledger months. The 5-day window with a 48-hour closing TWAP follows entry 9363. The asymmetric thresholds follow MetaDAO's documented template (non-team-sponsored proposals pass at +3%, team-sponsored at −3%, TWAP recording starts after a 24-hour delay) **FACT, docs.metadao.fi/governance/twaps, verified 8 Nov 2026**, but stay conservative: nothing passes when the market says NCV would get worse.
- **Quorum.** At least 12 unique traders and $6,000 of open interest, at most 50% of it subsidised (a design target: §5.2 shows today's depth is far below it). Below quorum, status quo and no re-vote for 30 days.
- **Caps and spam price.** No wallet above 10% of open interest; $200 proposal bond, refunded if the market clears quorum.
- **Execution.** On-chain actions (parameter change, envelope move) execute through the single new contract (§5.9); off-chain actions execute by steward order within 7 days with proof published, or the action reverts to a re-vote.
- **Steward veto.** Only for safety or legal reasons, with published reasons within 24 hours and an automatic ratification market within 7 days; a veto without ratification reverts.

5.6 Worked examples, including the market saying no to the guild itself

The format, a market that rejects the operators' own proposal and accepts an outsider's, is the one entry 9363 uses; the numbers are this plan's. Both are Month 10 fast decisions resolving on trailing NCV_{30}, which at Month 10 is $17,112 (the 30-day figure, not the 90-day NCV_{90} of $33,636) **ESTIMATE**.

Example 1, the market rejects the stewards' own proposal. The stewards propose raising the take from 10% to 12% "to accelerate the treasury". PASS (NCV_{30} if adopted) TWAPs at $16,753 (−2.1%); FAIL TWAPs at $17,112. PASS is below FAIL, so the fee rise would lose volume faster than it gains margin, and the proposal is rejected automatically. The market must be able to say no to its own operators.

Example 2, the market accepts a poster's proposal, at a capped size. A poster proposes a fifth category, "AI eval data labelling", with a $4,000 launch envelope. PASS TWAPs at $17,899 (+4.6%) against FAIL $17,112 with $8,400 of open interest. The adoption margin is cleared ($787 against a required 3% × $17,112 = $513), but the depth ladder limits binding spend at the market's measured depth; the first tranche is released at the cap and the remainder goes to a second market 30 days later. The category opens T0-only for its first 200 tasks.

5.7 The cold-start problem: a derived subsidy

Most plans name a subsidy; this one derives it, and the derivation is the deliverable.

Who trades, and why. Five populations: Seer power users and forecasters (about 16 per decision, drawn by tournament prizes), assessor agents (6), poster-hedgers (4, who buy FAIL as insurance and expect to lose), DAO treasury watchers (2), and the guild as subsidised market maker (an expected cost of $282 per decision).

The derivation ASSUMPTION**.** A trader who is right 60% of the time against a 50% market expects a 20% edge at fixed odds ($24 on a $120 median stake); in an LMSR with b = $850 the price moves against the trader and the expected profit falls to about $15. The plan budgets $18 per decision, the conservative end. With about 24 profit-seeking traders per decision the profit pool the market must pay is 24 × $18 = $432; poster-hedgers contribute about $150 in expected losses; **the residual the guild funds is $282 per decision**.

The market maker's worst-case loss per binary market is b × ln 2 = **$589.18**; expected realised cost is about 48% of that. The subsidy tapers: $350 per decision in Months 1–3, $300 in 4–6, $250 in 7–9, $200 in 10–12. The taper assumes poster-hedger and organic participation grow so that the residual falls; if they do not, the stop rule applies and b is reduced, which lowers the depth-ladder cap in proportion. At 30 Class A/B decisions a year the expected LMSR cost is about $8,250. The stop rule: when 3 consecutive decisions clear quorum with subsidised volume below 30% of open interest, seeding stops for Class B, then for Class A after 3 more.

COMPONENT	M1–3 /MO	M4–6	M7–9	M10–12	YEAR
LMSR seeding (expected)	$875	$750	$625	$500	$8,250
Tier-B batch-market seeding	$400	$350	$300	$200	$3,750
Forecaster tournament prizes (incl. random-slice budget)	$600	$550	$500	$400	$6,150
Worst-case buffer (unspent rolls to reserve)	$525	$350	$175	$100	$3,450
Envelope	**$2,400**	**$2,000**	**$1,600**	**$1,200**	**$21,600**

Year-1 governance cost is $21,600, 8.1% of gross revenue, inside the $11,500 average fixed cost and the break-even.

Where it could be wrong, now with data. The 24-trader and $120-stake assumptions are the load-bearing ones, and §5.2 shows today's organic depth is far below them. If fewer than 12 unique traders show up per market for three markets, quorum fails, decisions fall back to status quo and the Governance Crier doubles tournament prizes once. The safe failure mode is no decision, and the depth ladder means a thin market cannot bind much even when it does clear.

5.8 Fallback ladder and circuit breaker for futarchy itself

1. **Thin market** (open interest below $6,000): status quo. No decision is better than a manipulated decision.
2. **Two consecutive failed quorums:** markets drop to advisory; every advisory call is scored on its Brier score.
3. **Three consecutive failed quorums:** governance reverts to the steward council for at most 30 days with a published recovery deadline.
4. **Seer or Gnosis outage over 48 hours:** all markets pause; if over 30 days, the pre-specified Base-native LMSR fallback is activated.

Circuit breaker. Futarchy can be wrong systematically, not just noisily. A review by the Institute for Exotic Science (Sep 2026) reports no head-to-head test showing decision markets beat committees, and notes that such decisions are mostly capital allocation rather than daily operations REPORTED. MetaDAO states that it has run 96 proposals for 14 organisations since November 2023 FACT as stated on docs.metadao.fi, checked 8 Nov 2026. So the plan measures its own calibration and can switch it off:

- **Brier gate:** if 3 consecutive binding decisions resolve against the market's prediction (realised move opposite by more than 2 points), authority drops one stage (Stage 3 to 2, Stage 2 to 1, Stage 1 to advisory-only for 30 days), and re-activating Stage 1 takes a Class A market. *The calibration gate follows entry 9362; the 3-strike rule is this plan's.*
- **Manipulation tripwire:** a single cluster above 40% of volume, or a >25% price move in the final hour, extends trading by 72 hours and doubles the subsidy once.
- **Sunset:** every binding delegation expires after 12 months unless renewed by a Class A market.

Staged authority. Stage 1 (from Day 76): advisory plus narrow binding, fee moves of ±2 points, binding only where the depth-ladder cap covers the market's estimated 30-day NCV effect (otherwise advisory), and spend limited by the depth ladder. Stage 2 (after 4 of the last 6 resolved proposals were accurate within 15%): bond, sampling and take-rate parameters inside their bands. Stage 3 (after at least 10 resolved with ≥70% accurate): automatic on-chain execution. *The staged-authority idea follows entry 9378.*

5.9 Implementation: one new contract, and mandates

Settlement lives on Base (USDC escrow, x402-compatible machine payments) **ASSUMPTION: no public statement confirms Foreguild's chain; Base is a proposed design choice**. Governance markets live on Gnosis via Seer. The two are joined by a Reality.eth question on NCV_{90}. **One new contract**, `ForeguildSettlement` (about 400 lines): `escrow`, `release`, `refund`, `slashBond`, `setTakeRate` and `setEnvelope` (callable only on a passed market, timelocked 48 hours), `pause` (guardian only), `publishVerdict`. Everything else is live Seer infrastructure, configured, not built. Value never sits on a bridge: only attestations cross, so a bridge failure delays governance and never strands poster funds.

Mandates. A passed proposal does not become a memo; it compiles into a fixed-schema object that caps what agents may spend and which work may be listed, enforced by the Router before any job is listed. *The mandate concept follows entry 9373; compiling a passed proposal into a mandate with fields such as `source_fip` and `harness_share_bps`, and showing it in a runnable demo, is also the pattern of entry 9385. FIP stands for Foreguild Improvement Proposal.* The schema is exactly: `mandate_id, source_fip, market_id, take_bps, allowed_categories, daily_spend_cap_usd, max_ticket_usd, harness_share_bps, expiry, status`. A job is listed only if its category is allowed, its amount is within `max_ticket_usd`, and today's spend plus its amount is within `daily_spend_cap_usd`; otherwise it is BLOCKED, not queued. A mandate is born only from a proposal whose status is passed; an expired mandate enforces nothing. The runnable demo (§12.3) exercises all of this.

6. Treasury and spending: envelopes, runway, freeze rules

6.1 Envelopes

Money moves only inside named envelopes. No envelope, no spend.

ENVELOPE	$/MONTH	OWNER	RULE	KILL CONDITION (AUTOMATIC)
Operating (fixed)	9,700	Treasury Office	Published weekly; any line +20% needs a Class B market	A line +20% without a passed market freezes that line
Governance subsidy	2,400 → 1,200	Governance Crier	§5.7; unspent rolls to reserve, never to ops	3 consecutive failed quorums stop LMSR seeding
Growth (acquisition)	800 → 5,000	Scout	§2.4; cut to $0 by Freeze Rule 1	Revenue <70% of plan for 2 months → $0 until recovery
Audit (capitalised)	$18,000 pre-launch + $6,000 in Month 9	Treasury Office	§8.5	No report by Day 42 → uncapped settlement does not start
Assurance reserve	Floor $5,000; 2× trailing-90-day claims	Treasury Office	§1.2	Reserve below floor → assurance sales pause

Capital to raise before Day 1 **ASSUMPTION**: **$104,700**: $18,000 for the pre-launch audit, $14,700 to fund Days 1–30 (one month of fixed cost, $9,700, plus $5,000 of legal and counsel review) and **$72,000 of opening cash on Day 31**, where the ledger starts. It comes from design-partner commitments, a grant or sponsor capital, and explicitly **not** from the contest prize. The base-case treasury trough is $27,530, only $2,530 above the $25,000 floor, so the raise has no slack: a raise smaller by more than $2,530 breaches the floor in Month 6. Year-end surplus split: 70% retained for growth, 20% to the governance reserve, 10% to contributor rewards.

6.2 Monthly ledger, Year 1 (base case)

Formulae **ASSUMPTION**: take = 10% × GMV; assurance gross = 1.8% × GMV (45% × 4%); assurance claims = 50% of gross; verification = $0.80 × tasks; gas and dispute = $0.15 × tasks; contribution = take + assurance net − verification − gas/dispute; fixed = $9,700 + subsidy; other = harness + console + API. Totals are computed from unrounded values. NCV_{30} is the month's surplus; NCV_{90} is the trailing three-month sum (provisional before Month 3). Every row below is produced by `model.py`.

MO	TASKS	AOV	GMV	TAKE	ASSUR. NET	VERIF.	CONTRIB.	OTHER	FIXED	GROWTH	SURPLUS (NCV_{30})	NCV_{90}	CUM.	CASH
1	400	$48	$19,200	$1,920	$173	$320	$1,713	$0	$12,100	$800	−$11,187	−$11,187*	−$11,187	$60,813

MO	TASKS	AOV	GMV	TAKE	ASSUR. NET	VERIF.	CONTRIB.	OTHER	FIXED	GROWTH	SURPLUS (NCV_{30})	NCV_{90}	CUM.	CASH
2	650	$50	$32,500	$3,250	$293	$520	$2,925	$0	$12,100	$1,000	−$10,175	−$21,362*	−$21,362	$50,638
3	900	$52	$46,800	$4,680	$421	$720	$4,246	$0	$12,100	$1,300	−$9,154	−$30,516	−$30,516	$41,484
4	1,200	$54	$64,800	$6,480	$583	$960	$5,923	$150	$11,700	$1,600	−$7,227	−$26,556	−$37,743	$34,257
5	1,600	$56	$89,600	$8,960	$806	$1,280	$8,246	$598	$11,700	$2,000	−$4,856	−$21,236	−$42,598	$29,402
6	2,100	$58	$121,800	$12,180	$1,096	$1,680	$11,281	$947	$11,700	$2,400	−$1,872	−$13,954	−$44,470	$27,530
7	2,600	$60	$156,000	$15,600	$1,404	$2,080	$14,534	$1,645	$11,300	$2,800	$2,079	−$4,648	−$42,391	$29,609
8	3,200	$61	$195,200	$19,520	$1,757	$2,560	$18,237	$2,144	$11,300	$3,200	$5,881	$6,088	−$36,510	$35,490
9	3,900	$62	$241,800	$24,180	$2,176	$3,120	$22,651	$2,892	$11,300	$3,600	$10,643	$18,603	−$25,867	$40,133
10	4,700	$64	$300,800	$30,080	$2,707	$3,760	$28,322	$3,690	$10,900	$4,000	$17,112	$33,636	−$8,755	$57,245
11	5,600	$66	$369,600	$36,960	$3,326	$4,480	$34,966	$4,439	$10,900	$4,500	$24,005	$51,761	$15,250	$81,250
12	6,600	$68	$448,800	$44,880	$4,039	$5,280	$42,649	$5,437	$10,900	$5,000	$32,186	$73,304	$47,437	$113,437

* NCV_{90} is provisional before Month 3.

Year-1 totals: **33,450 tasks · GMV $2,086,900 · gross revenue $268,196 · operating surplus $47,437 · surplus after audit $23,437 · closing cash $113,437**. Treasury trough $27,530 at the end of Month 6.

6.3 Scenarios, each with the rule that fires

The model applies Freeze Rule 1 only where a row says so and never applies Freeze Rule 2, so cash troughs below $25,000 are shown as they would be without that second freeze.

SCENARIO [ASSUMPTION INPUTS]	TASKS	GMV	OPERATING SURPLUS	CASH TROUGH	MONTHS BELOW $25K FLOOR	CLOSING CASH	WHAT FIRES
Base	33,450	$2,086,900	$47,437	$27,530 (M6)	0	$113,437	—
Bear (volume ×0.6, AOV ×0.85, other ×0.5), no intervention	20,070	$1,064,415	−$62,274	−$9,441 (M9)	8	$3,726	Cash goes negative
Bear with Freeze Rule 1 firing in Month 3	20,070	$1,064,415	$526	$30,285 (M7)	0	$66,526	Freeze Rule 1 (§6.4)
Bull (volume ×1.4, AOV ×1.10, other ×1.5)	46,830	$3,213,826	$168,532	$42,747 (M4)	0	$234,532	Surplus split; growth envelope +30% by Class B market
Zero harness revenue	33,450	$2,086,900	$37,437	$26,580 (M6)	0	$103,437	Nothing; first surplus still Month 7
Assurance attach 20%	33,450	$2,086,900	$37,002	$25,656 (M6)	0	$103,002	Break-even slips to 2,174 tasks a month
Ex-Seer (no Seer-sourced tasks)	26,760	$1,669,520	$11,446	$21,714 (M7)	4	$77,446	Freeze Rule 2: frozen, not dead (§10.4)
Lean: 20% of base volume and non-take revenue, Freeze Rule 1 in Month 3	6,690	$417,380	−$63,873	$410 (M11)	7	$2,127	Month-3 kill signal at Day 120; raise nearly exhausted without it

Cash runway by scenario



33,450 Year-1 tasks	**$2,086,900** GMV	**$47,437** Operating surplus	**$113,437** Closing cash
$27,530 (M6) Cash trough	**Month 7** First surplus month		

- **Bear** (volume ×0.6, AOV ×0.85, other revenue ×0.5): without intervention, cash goes negative. The freeze rule below is why the bear case survives: bear revenue runs at about half of plan from Month 1, so Freeze Rule 1 (revenue below 70% of plan for 2 consecutive months) fires in **Month 3**. With the freeze the bear case ends Year 1 at an operating surplus of $526, a cash trough of $30,285 (Month 7) and closing cash of $66,526; the price is that growth spend stops for the rest of the year. *Version 1 printed different bear and bull figures that cannot be reproduced from its own stated inputs; the figures here come from* `model.py` *and supersede them.*
- **Lean (committed-demand) case.** Version 1 ramps to 6,600 tasks a month. The other plans in this contest assume roughly 900 to 4,000. The lean case runs the same model at **20% of base volume and 20% of non-take revenue** (harness, console and API scale with volume; about 1,300 tasks a month by Month 12, 6,690 in Year 1) with Freeze Rule 1 firing in Month 3. It is the plan's real downside, and it is not survivable: the operating surplus is −$63,873, the raise is almost exhausted (cash trough $410 in Month 11, closing $2,127) and cash is below the $25,000 floor in 7 months. The plan does not let it run that far. At Day 120 the Month-3 kill signal fires (cumulative GMV $19,700 against a $50,000 threshold), when cash is still $39,277 of the $104,700 raise; the stewards put a continue-or-wind-down proposal to a market, and a wind-down returns escrowed funds and keeps what is left. Bounded downside, not a viable plan at that demand, is the honest reading. (The model does not apply Freeze Rule 2, which would cut spending further, so its cash figures understate cash.)
- **Zero harness.** If harness royalties are $0 all year, operating surplus is $37,437, break-even stays in Month 7 and the cash trough is $26,580 (above the floor by $1,580). The most speculative line cannot sink the plan.
- **Assurance attach 20%** (down from 45%): contribution per task falls from $5.59 to $5.29, break-even rises from 2,057 to 2,174 tasks a month, Year-1 surplus falls by $10,435 and the cash trough is $25,656, only $656 above the floor.
- **Ex-Seer** (no Seer-sourced tasks: volume ×0.8, other ×0.85, growth ×0.8): cash is below the $25,000 floor in 4 months (trough $21,714), so Freeze Rule 2 applies and the guild runs frozen rather than dead (§10.4).

6.4 Freeze rules: automatic, not discretionary

- **Rule 1:** revenue below 70% of plan for 2 consecutive months → growth envelope to $0, subsidy and steward draw halved, until revenue recovers above 85% of plan for one month (*structure and thresholds follow entry 9363*).
- **Rule 2:** treasury below $25,000 → all discretionary spend stops except escrow obligations and the assurance reserve. Only a Class A market can authorise spending below the floor.
- **Rule 3:** acceptance-without-dispute below 92% in 30 days → parameter-change markets suspend.

- **Rule 4:** blended verification above $3.00 a task for 2 weeks → T2 sampling to 5% and a fee market opens.

Freeze rules execute by contract and by published procedure. They do not wait for a meeting.

7. Disputes and human intervention: the ladder, the guardrails, the kill criteria

7.1 Dispute ladder (clocked)

STEP	WHAT HAPPENS	CLOCK	COST / BOND
L0, auto re-check	Delivery re-run through T0 and T1 with fresh judges	0–2 h	Free
L1, senior reviewer	One senior bonded reviewer plus a rubric re-score	24 h	Challenger bond $25; loser pays
L2, human panel	3 bonded humans, majority rules, reasons published	72 h	Bond $100; loser pays
L3, Reality.eth + Kleros	Tasks over $1,000 or precedent-setting cases	Per Kleros	Bond $250; the Kleros ruling is canonical

Target: dispute rate below 3%, median L1 resolution under 24 hours. Poster silence counts as acceptance after the 24-hour challenge window, because verification, not silence, is the acceptance mechanism.

7.2 Human intervention: three powers, nothing else

Stewards (2 part-time humans) hold exactly three *emergency* powers: pause the escrow contract on suspected exploit, veto a passed market for safety or legal reasons (subject to ratification, §5.5), and appoint T2 panel members. Their ordinary duties are procedural: they may sponsor proposals (steward-sponsored Class A proposals face the lower 2% margin), execute off-chain actions of passed markets by order, and take Class C decisions below $1,000 inside an envelope. They cannot spend beyond that, set fees outside a market, or decide disputes outside the ladder. Every emergency action faces a ratification market within 7 days or reverts automatically. In the fallback ladder (§5.8, step 3) the stewards act as a caretaker council for at most 30 days, limited to Class C decisions and the emergency powers; they cannot adopt Class A or B changes. A guardian multisig (3-of-5, including one non-Seer member) can pause but never spend. Guardian sunset: after NCV_{90} has been positive for 120 consecutive days the guardian may pause only for security incidents. *The sunset idea follows entry 9385.*

7.3 Kill criteria: written before launch

CRITERION	TRIGGER	ACTION
Seer dependence	Seer-sourced tasks above 35% of GMV at Day 90	Wind down Seer-sourced tasks to ≤25% within 30 days; if impossible, pause new listings
Dispute spiral	Dispute rate above 8% in 30 days	Halt matching in the affected category; Verification Economics Office review
Verification blowout	Blended verification above $6.39 a task (contribution zero at AOV $60)	Halt new listings; emergency Class A market on take rate and listing rule
Futarchy failure	3 consecutive binding decisions resolve against prediction	One stage down (Stage 1 falls to advisory-only for 30 days)
Treasury	Cash below the $25,000 floor	Freeze Rule 2
Quality	Acceptance-without-dispute below 85% in 30 days	Halt new categories; full T2 audit of the last 500 tasks

7.4 What happens if the thesis is wrong

If realised AOV stays below $40 for two months, the guild stops listing data labelling, reprices the take by Class A market and halves the growth envelope. If NCV_{90} is negative after Month 9, the stewards must put a wind-down plan to a Class A

market: return of escrowed funds, orderly closure of open markets and publication of the full ledger. A plan that cannot describe its own ending is not a plan.

8. Risks and mitigations

8.1 Risk table

RISK	LIKELIHOOD / IMPACT [JUDGEMENT]	MITIGATION	OWNER	MONTHLY METRIC
Seer dependency: futarchy stack slow, down or changed	Medium / High	§5.8 fallback ladder; pre-written Base-native LMSR; governance never strands settlement	Governance Crier	% of decisions clearing quorum; Seer share of GMV
Thin futarchy markets and manipulation	**High / High**	§5.2 measured depth; §5.3 depth ladder; quorum; 10% wallet cap; manipulation tripwire; advisory-first	Governance Crier	Unique traders per market; capital-to-move vs bound spend
AI judge bias, correlated model failure	High / High	Two model families, calibration scoring, 10% human sample, precedent library	Verification Economics	T2 overturn rate; blended verification $/task
Sybil and reputation gaming	High / Medium	Reputation only from settled escrow; unique-counterparty weighting; bonds; probation lane; recycled-worker netting	Dispute & Safety	Flagged pairs; wash volume excluded
Prompt injection via listings, messages, deliverables	High / Medium	All text is untrusted input: sandboxed execution, no secrets in agent context, output schema validation	Dispute & Safety	Injection incidents caught per month
Key and fund security	Medium / Critical	Non-custodial escrow contract (never a house wallet); keys never in agent context; executor separation; per-order caps	Treasury Office	Unreconciled escrow days (target zero)
Verification cost inflation	Medium / High	§4.4 grids; $3.00 tripwire; governed T2 sampling	Verification Economics	Blended $/task vs $3.00
Regulatory: escrow and custody classification	Medium / High	Non-custodial design; counsel review before launch ($5,000 legal line); no fiat; jurisdiction review is a Day-30 gate	Treasury & Legal	Counsel sign-off
Demand shortfall	Medium / High	Bear and lean cases with freeze rules; ex-Seer survival (§10.4)	Scout	Settled tasks vs plan; CAC by channel
Bridge and multi-chain	Low / High	Only attestations cross; value stays put	Treasury Office	Bridge downtime

8.2 Dependency on Seer: the honest paragraph

Foreguild's governance runs on Seer's stack, and Seer is also a customer (20% of tasks in the base plan), the prize sponsor (4,000 USDC), the selection mechanism (every plan becomes an outcome in a Seer market and the Seer team makes the final choice) and a competitor in this contest (Seer team members participate) **FACT, bounty 1531**. Four distinct dependencies on one counterparty. The plan prices each: governance has a fallback ladder (§5.8); the customer relationship is capped and kill-switched (§7.3); the prize is not assumed as float (§6.1); the contest dynamics are addressed in §10.

8.3 Data ownership and consent

Task content belongs to the poster; delivery output is licensed to the poster on acceptance. The guild stores specs, checks and verdicts for 90 days, then archives hashes only. Eval-data licensing happens only with explicit opt-in from both poster and worker, with 50% of the revenue shared with the contributors. No consent, no sale, no exception.

8.4 Human staffing: who actually runs this

Beyond the 2 stewards ($4,800 a month), the four fixed-cost offices are run by agents and part-time humans with the human hours costed inside the $9,700 envelope. Total fixed operating cost in Year 1 is $116,400, of which human cash cost is about $109,200. The plan does not pretend agents run themselves for free.

8.5 Smart-contract audit: budgeted, scoped, timed

Budget **$24,000** ASSUMPTION: $18,000 pre-launch audit of `ForeguildSettlement` (the only new contract) plus a $6,000 re-audit in Month 9. Timeline: scoping Days 15–21, code freeze Day 28, fieldwork Days 29–42, report Day 42. The first settlements (Days 31–45) run on a capped $5,000 pilot escrow until the report lands; uncapped settlement starts Day 46. Out of scope: Seer's own contracts.

8.6 Why there is no token

(1) The welfare metric is NCV_{90} in USDC, and a token would force it to be priced in a thin, manipulable asset. (2) Decision markets on a token price measure speculation, not guild health. (3) A token creates a second constituency whose interests diverge from posters and workers. (4) No distribution is fair on day one. A token may be proposed by a Class A market after Month 12, only with a full design attached.

9. Roadmap: 90 days, measurable milestones

Day 0 = adoption (13 November 2026). Operating months in the ledger begin when escrow goes live (Day 31).

WINDOW	WORK	EXIT GATE (A NUMBER, NOT AN ADJECTIVE)
Days 1–14, Spec	Finalise specs for 4 categories; write `ForeguildSettlement`; sign 20 design-partner LOIs; open forecaster-tournament registration; publish the 12 named templates	20 LOIs signed; spec-linter live; contract code-complete
Days 15–30, Build and audit	Audit scoping and code freeze Day 28; Seer market templates configured; T0 harness library (30 checks); dispute ladder drilled on testnet; **depth measured weekly and the ladder set**	Audit engaged; 30 T0 harnesses passing; pilot escrow capped at $5,000
Days 31–45, First settlements	Paid pilot on Base, escrow capped at $5,000, first 25 funded tasks from design partners; T1 judges calibrated; NCV published weekly from Day 38	≥25 settled in total AND ≥10 from external buyers AND the external set's contribution ≥$0 AND dispute rate <15%
Days 46–60, Open board	Public listings in 2 categories; Router with Wilson scoring live; first batch assessment market	≥400 settled (Month 1 ledger); ≥10 repeat posters; verification ≤$1.00 a task
Days 61–75, Futarchy on (Month 2)	First advisory markets (they resolve 30 days after adoption); tournament pays first prizes; categories 3–4 open; console subscriptions on sale; first random slice	≥3 advisory markets opened; ≥12 unique traders per market; Month-2 ledger 650 tasks
Days 76–90, First binding decision	Stage 1 opens: the first binding decision is *adopted* within the depth-ladder cap (it resolves at least 30 days later); a Verified badge for categories that pass the panel gate; re-audit scoped; Seer-share check	≥1,050 cumulative settled and ≥$51,700 GMV (Months 1 and 2 of the ledger); first binding decision adopted; Seer ≤35% of GMV; ≥55 of the first 100 posters with a second task; ≥60% human-posted
Days 91–120, Month 3	Stage 1 decisions resolve; Brier gate scored; Stage 2 considered	≥1,950 cumulative settled and ≥$98,500 GMV; first binding decision resolved

Day-90 scoreboard ASSUMPTION targets**:** 1,050 cumulative settled tasks · GMV $51,700 · dispute below 5% · verification ≤$1.00 a task · NCV published weekly (8×) · ≥12 unique traders per market · Seer ≤35% of GMV · treasury ≥$40,000 (plan: $50,638). Missing any two gates triggers the freeze rules (§6.4) for the next quarter's envelopes.

Month-3 kill signal (Day 120). If cumulative GMV is below $50,000 or settled tasks below 1,000 on Day 120, against a plan of $98,500 and 1,950, the stewards must put a continue-or-wind-down proposal to a Class A market within 7 days. *A Month-3 kill signal also appears in entry 9378; the thresholds here are re-derived from this plan's own ledger.*

10. Strategy with internal Seer submissions in the field

This section exists because the contest's structure demands it. Every plan becomes an outcome in a Seer market, the Seer team makes the final choice, and Seer team members participate **FACT, bounty 1531**. The jury competes. Pretending otherwise would be the least credible thing in this document. (As far as this plan could verify, no other entry addresses it beyond labelling itself "non-internal" or describing the contest's structure.)

10.1 The game, stated plainly

Traders in the selection market price two things: which plan is best, and which plan the Seer team will pick. The gap between them is where an internal submission has an edge: traders rationally discount any external plan by the probability that the team prefers its own. This plan does not try to close that gap with rhetoric. It changes the payoff of picking it through three structural moves.

10.2 The three moves

1. **Verifiability is priced; private information is not.** An internal plan can rest on data only Seer holds (backlog size, trader counts, real costs). Traders cannot verify those claims, so they must discount them. This plan rests on nothing private: every external fact is cited and re-checked (§12.2), every assumption is labelled with its failure sign (Appendix A), and the headline numbers are recomputed by a public model (`model.py --check` pins 48 of them). A trader can re-price this plan in an afternoon.
2. **Adoptability beats authorship.** The plan is written so Seer can adopt it on 13 November without rewriting it: one new contract, named offices with budgets, a ledger, freeze rules and Day-1 decisions in Appendix C. An internal plan may rest on knowledge outsiders cannot check; an external plan has to be executable by someone who was not in the room.
3. **Alignment without capture.** Every governance cycle in this plan is Seer trading volume, but the guild earns nothing from it (§5.4). The plan caps Seer-sourced GMV at 25% in Year 1 (20% in the base plan) with a kill criterion at 35%, so that adopting it does not depend on Seer's own backlog.

10.3 What the insider knows that this plan does not

The insider knows Seer's true backlog, trader base and costs. This plan answers with two hedges: the subsidy derivation (§5.7) is built to survive its assumptions being wrong by 2× (12 traders instead of 24 means advisory fallback, not collapse), and the economics do not require the insider's assets at all (§10.4). Where an insider plan has real demand data, this plan offers measured market-depth data of its own (§5.2) and a repeat-poster gate.

10.4 The ex-Seer test: the plan survives its own sponsor

Remove the entire Seer channel (20% of tasks; other revenue ×0.85; growth ×0.8) **ESTIMATE, same model**: Year-1 tasks fall to 26,760, operating surplus to $11,446, break-even slips from Month 7 to Month 8, closing cash $77,446. The plan is thinner and later. Cash dips below the $25,000 floor in 4 months (trough $21,714), so Freeze Rule 2 applies and the guild runs frozen, not dead. A plan that dies without its sponsor's demand is not an operating plan; it is a grant application.

10.5 If an internal plan wins

The mechanisms are modular and attributed: the verification P&L (§4.2), the depth ladder (§5.3), the derived subsidy (§5.7) and the futarchy circuit breaker (§5.8) are each adoptable independently, and this document says so. A plan that can only win whole is a plan that cannot be partially right.

11. The field: how this plan compares with the other entries

When fetched on 7–8 November 2026 the bounty page listed **48 claims**. This plan downloaded every plan it could open (38 of 48 were readable in full or in substantial part; ten were not: 9358, 9367, 9380, 9382, 9388, 9394, 9395, 9431, 9447 and 9455 are cloud-drive login pages, a blocked gateway, an installer link, an image only, or a description with no document) and read them. The table lists the entries that are strongest on a dimension this plan also cares about. Everything is as read from the entry; none of it was re-run unless §12 says so. The point of the table is to say plainly where others are ahead.

ENTRY	STRONGEST ON	WHERE THIS PLAN DIFFERS
9414	A runnable 24-month model with checks and a simulator; sensitivity solver; measured Seer market list (548 conditional markets); thin-market rule; random slice; tranche and hard-stop discipline; its own committed-volume and downside cases	This plan's base case reaches a Year-1 surplus (an assumption-driven figure) where 9414's base case loses money in Year 1; this plan adds a derived subsidy and the ex-Seer test, and adopts 9414's random slice and panel gate
9396	Live Seer markets on Base, a self-funded Verdikta task with payout evidence, an interactive calculator, a near-identical verification thesis	No derived subsidy, sensitivity grid or monthly ledger in 9396; this plan has no live market of its own
9385	Runnable mandate demo, a live Seer decision market (Gnosis), SEC take rates, joint-bond Guilds	No 12-month ledger or derived liquidity in 9385; this plan answers the Guild model in §3.5 and rebuilds a runnable demo
9442	Runnable model, measured pool depth, stake-cap argument, explicit kill rules	This plan adds a ledger, a verification P&L and the depth ladder built on that stake-cap idea (credited)
9461	Demand-first managed desk, human-vs-agent revenue split, graduation from advisory to binding, five written first markets	This plan adopts the split and the market-worthiness test (credited) and adds a ledger, sensitivity grids and a derived subsidy
9363	A full 12-month ledger with nine offices	No live proof or derivation of liquidity in 9363; this plan matches the ledger and adds the verification and subsidy derivations
9477	Primary Verdikta data (394 bounties), reproducible scripts, a rubric-defect back-test, checkable day-90 predictions	This plan has no comparable primary data on job boards; it adopts netting of recycled volume (credited)
9424	Fully written Reality.eth question, noise and cost tests, 39 sources	Scripts in 9424 were placeholders at the time of reading
9378	Per-task credit markets, accuracy-based authority, payer caps, revenue from week 1	Staged authority and the dual horizon are adopted here (credited); no 12-month ledger in 9378
9436	Parent-normalised scalar markets with a conservative interval, executable model with tests	Adopted here (credited); no verification P&L or ledger in 9436
9381	A runnable demo that fixes a real bug, with 13 browser checks and a candid pilot plan that does not break even	Standing orders adopted here (credited)
9465	Primary poidh data (105 bounties, 1,381 entries) and a contest-desk revenue model	No ledger of comparable depth in 9465; this plan did not collect poidh data of its own
9408, 9423, 9425	Live Seer parent and conditional markets on Base, a self-funded Verdikta task	Market liquidity is near zero by the authors' own statement; this plan has no live market

What this plan has that, as far as it could tell, no other entry combines: a derived liquidity subsidy (LMSR b = $850, $282 expected cost a decision, a taper and a stop rule) *and* a 12-month ledger *and* a verification P&L with a two-way sensitivity grid *and* a budgeted contract audit *and* a written strategy for the fact that the jury also competes *and* an ex-Seer survival test *and* a lean case. Several entries have one or two of these; none among the 38 readable entries was seen with all of them.

What this plan lacks that others have: primary data on job boards or markets (9477, 9462, 9465, 9414, 9424, 9442), and a live on-chain market (9385, 9396, 9408, 9423, 9425). Several of those live markets and the Verdikta task behind them are self-funded and nearly empty, by their authors' own statement. Section 12 shows what this plan measured itself, and the entry kit states what it did not.

Credits. Mechanisms adopted from other entries, marked where it is used or, for structural patterns, listed here: **9363**: the structure and thresholds of Freeze Rule 1, the 92% acceptance floor, the 55-of-100 repeat gate, the 5-day window with a 48-hour closing TWAP, the T2 sampling pattern and the format of a market rejecting the operators' own proposal. **9374**: the 6% check-cost listing rule, the 70/30 harness split, CMV_{90} as the ancestor of NCV_{90}, the T0/T1/T2 tier names with a deterministic first tier, the A/B/C decision tiers and the score-based auction frame. **9385**: the compile-a-passed-proposal-into-a-mandate pattern and field names, the guardian sunset, and the idea of a demo that runs. **9378**: the dual horizon, tiered payer caps, staged authority and a Month-3 kill signal. **9442**: bounding authority by manipulation cost. **9414**: random

slice and panel gate. **9461**: human-versus-agent reporting and the market-worthiness test. **9477**: netting of recycled worker volume. **9436**: parent-normalised prices. **9381**: standing orders. **9362**: the calibration gate. **9373**: the mandate concept. The demo's fixtures and numbers are this plan's own; its three-part structure (harness, compiler, enforcer) follows the pattern of entry 9385.

12. Evidence: what was measured, what was verified, what can be run

12.1 Measured by this plan

WHAT	HOW	WHERE
Seer market depth: counts, traded volume, liquidity, open interest, by chain and type	`python3 scripts/market_depth.py` reads Seer's public `markets-search` function page by page (3,051 markets in the default view, 1,153 in the conditional view) and writes the raw JSON plus a summary	`data/market_depth.json`, raw `.json.gz`
Cost to move a Seer-style market	`model.py` (`attack_cost_3pct`, `lmsr_move_cost`), closed-form LMSR	§5.3
The figures quoted in §1 to §6 and §10	`python3 -I model.py --check` verifies 48 pinned figures (19 from version 1 and 29 new in version 2); four version-1 scenario figures that do not reproduce are listed with the model values	`model.py`, `test_model.py`

12.2 External facts, re-checked against the source on 8 November 2026

CLAIM USED IN THE PLAN	RESULT
Fiverr FY2025: marketplace revenue $297.5M, marketplace GMV $1,073.0M, take rate 27.7%, total revenue $430.9M (Form 20-F)	Matches the filing
Upwork FY2025: Marketplace revenue $682.9M; reported Marketplace take rate 18.7% (Marketplace GSV); total GSV $4,028M	**Corrected from version 1**: total GSV is not Marketplace GSV; dividing by it gives 16.9%, not 18.7%
MetaDAO: non-team proposals pass at +3%, team-sponsored at −3%, TWAP recording after a 24-hour delay; 96 proposals for 14 organisations	Matches docs.metadao.fi/governance/twaps and /governance/overview (organisation's own count)
Olas Mech: 2,103,560 requests, 100 mechs, 15% fee, about $104k turnover; 97.2% completion	Matches the Blokz article (the completion rate is the article's own calculation)
Gitcoin GG24: $350,000 allocated via deep funding	Matches the Gitcoin governance forum; $350,000 is the deep-funding allocation for one category, not all of GG24
Seer's dedicated futarchy contracts are documented on Gnosis; core contracts on Gnosis, Ethereum, Optimism and Base	Matches the deployed-contracts page
Seer's `markets-search` function and its parameters	Matches the documentation at docs.seer.pm/developers/api
Reported take rates (Upwork 0–15%, Fiverr 20% plus 5.5%, Freelancer.com 10% or $5), the July 2026 Kleros experiment, the Institute for Exotic Science review	Not verified from a primary source here; labelled REPORTED
The contest announcement and the identity of the naming-contest winner	Could not be verified from a primary source; the plan does not rely on either

Figures taken from other entries (for example the rubric back-test in §4.1, which entry 9477 itself describes as a consistency check on the same history) are labelled as reported by those entries and were not re-run here.

12.3 Runnable artefacts

- **`model.py`**: one model behind the ledger, scenarios, unit economics and depth-ladder numbers; `--check` exits 0 only if every pinned figure matches.

- `demo/run_demo.py` : a deterministic local demo, no network, no funds. It shows (1) a T0 harness that rejects a bad dataset-dedup task and accepts a good one, (2) a mandate compiler that turns a *passed* proposal into the fixed schema of §5.9 and refuses a proposal that has not passed, and (3) an enforcer that allows or blocks four jobs against the active mandate (category, daily cap, ticket cap). Its output is committed in `demo/captured-output.txt` .
- **This page**: the calculators, the ledger and the mandate demo run in the browser from the same formulas; a parity test compares the browser code with `model.py` .

What none of this proves. The demo is not an on-chain contract and moves no funds; it is not T1 or T2 verification; the model's inputs are assumptions, not observed demand. The depth measurement is a snapshot of one public API on one day.

12.4 The author's own bounty record: ten real claims

Before proposing a market for verified work, the author worked the demand side of the existing ones. Pools, rival-claim counts and outcomes below were read from the poidh API on 8 November 2026 MEASURED; claim numbers can be looked up on poidh, and the table is also in `data/own_bounty_record_2026-11-08.json` . USD values use poidh's quoted ETH price of that day.

BOUNTY	CLAIM	POOL	RIVAL CLAIMS ON IT	STATE ON 8 NOV	POOL ÷ CLAIMS
base/1512	9306	$4.47	15	decided, **lost** (winner 9162)	$0.30
base/1513	9307	$2.23	17	decided, **lost** (winner 9131)	$0.13
arbitrum/339	1160	$2.23	12	decided, **won** (0.000975 ETH credited)	$0.19
base/1510	9308	$4.47	22	open, no decision after 13 days	$0.20
base/1515	9309	$33.50	25	open, 11 days	$1.34
base/1520	9310	$35.73	28	open, 9 days	$1.28
base/1525	9311	$60.97	8	**cancelled by the poster** after the claim was filed; no winner	$7.62
base/1519	9315	$15.41	10	open, 9 days	$1.54
base/1093	9317	$27.92	6	open	$4.65
mainnet/45	401	$502.47	2	open, 4 days	$251.24

Ten bounties, $689 of pools, 14.5 rival claims per bounty on average. Of the four that have closed, one paid the author, two paid someone else and one was cancelled with the work delivered. Six are undecided, four of them after nine days or more.

What this does and does not show. It supports four design choices of this plan: (1) with 6 to 28 claims on pools of $2 to $61, the pool divided by the number of claims is $0.13 to $7.62, which is why §3 prices expected cost of failure and routes by score, not by who is first; (2) a poster can cancel after delivery (base/1525), which is why escrow is locked at posting and acceptance follows the harness and the dispute ladder, not the poster's discretion (§7.1); (3) four undecided bounties have waited 9 to 13 days, which is why the Priority tier buys latency (§3.4) and the first public board is limited to categories whose acceptance can be checked by a harness; (4) the one large pool in the sample ($502) drew only 2 claims, so a large pool alone does not guarantee a crowded field; one observation, no cause is claimed. It does **not** show Foreguild revenue, a take rate, or demand from paying buyers: these bounties are posted by others, the sample is ten, and the win rate (1 of 3 decided) is too small to estimate anything. Compute and time cost per claim were not tabulated. Entry 9457 publishes a task-level ledger with hours and net per task; this table reports outcomes only.

Appendix A: facts, assumptions and how each could be wrong

ASSUMPTION	VALUE	WRONG IF... (EARLY-WARNING SIGN)
AOV ramp $48 → $68	Year-1 average $62.39	Realised AOV <$40 for 2 months → §4.4 mitigations fire

ASSUMPTION	VALUE	WRONG IF... (EARLY-WARNING SIGN)
Take rate 10%	Band 7–14%	The price experiment (§1.4) shows volume elasticity above 1 → Class A market
Verification $0.80 a task	Mix 56/34/10	Blended cost >$3.00 for 2 weeks → §4.4 tripwire
24 active traders, $120 median stake	§5.7	<12 unique traders per market for 3 markets → tournament doubles; **today's measured depth is far lower (§5.2), which is why authority is capped by depth**
Blended CAC $18.35	1,755 posters	Agent-channel CAC >$15 → blended CAC $22.20, LTV/CAC 5.0× (3.5× if only 70% of tasks are attributed to acquired posters)
Capital raise $104,700	$72,000 opening cash on Day 31	A raise smaller by more than $2,530 breaches the $25,000 floor in Month 6 (the base trough is $27,530)
Harness royalties $10,000	about 33,000 checks	Zero harness → surplus $37,437, break-even still Month 7
Assurance attach 45%, loss 50%	Net +$0.54 a task	Loss ratio >65% in 30 days → repricing market
Volume ramp 400 → 6,600 tasks a month	33,450 in Year 1	Realised volume at a fifth of plan → the lean case (§6.3)
75% of tasks human-posted	channel mix	Below 60% at Day 90 fails the Day-90 gate; below 50% for two cycles raises the revenue-honesty alarm

What this plan does not know: who won the Foreguild naming contest; whether Seer is formally Foreguild's sponsor; whether Foreguild will run on Base (proposed, not stated); the true size of Seer's own backlog.

Appendix B: sources

1. Bounty 1531, https://poidh.xyz/base/bounty/1531 (API: /base/bounty/1531/data)
2. Seer documentation, https://docs.seer.pm and https://seer-3.gitbook.io/seer-documentation (deployed contracts, futarchy and conditional markets, `markets-search` API)
3. Hanson (2013), "Shall We Vote on Values, But Bet on Beliefs?", doi:10.1111/jopp.12008
4. MetaDAO, https://docs.metadao.fi/governance/twaps and /governance/overview
5. Kleros Foresight, https://docs.kleros.io/products/foresight
6. Olas Mech Marketplace, https://www.blokz.dev/articles/agents-hiring-agents-mech-marketplace
7. Gitcoin GG24 deep funding, https://gov.gitcoin.co/t/deep-funding-gg24-web3-tooling-and-infra-round/25040
8. Fiverr FY2025 Form 20-F, https://www.sec.gov/Archives/edgar/data/1762301/000117891326000858/zk2634486.htm
9. Upwork FY2025 Form 10-K, https://www.sec.gov/Archives/edgar/data/1627475/000162747526000012/upwk-20251231.htm (Marketplace revenue $682.9M; take rate 18.7% on Marketplace GSV; total GSV $4,028M)
10. Institute for Exotic Science, futarchy and decision markets in governance (Sep 2026)
11. Other contest entries cited by number in the text, as filed on poidh base/1531
12. ERC-8004, x402, BasedAgents and agent-risk reports: secondary sources, used only as indicative and labelled where used

Appendix C: decisions Seer must make on Day 1

1. Adopt NCV_{90} as defined in §5.1, including exclusions and the change rule, or state what changes.
2. Fund the $104,700 raise (audit $18,000, Days 1–30 $14,700, opening cash $72,000) and name the source; the contest prize is not part of it.
3. Appoint the 2 stewards and the 3-of-5 guardian multisig (one non-Seer member).
4. Confirm the chain split (settlement on Base, governance on Gnosis via Seer) or direct an alternative before build starts.
5. Approve the $24,000 audit budget and the code-freeze date (Day 28).
6. Decide whether to hold a reserve above the plan: the ex-Seer case dips below the $25,000 floor by about $3,300.
7. Set the Day-90 scoreboard (§9) as the first review point, and the consequence of missing two gates.
8. Approve the depth-ladder rule (§5.3) and the weekly depth measurement that feeds it.

Launch checklist (each item signed off by its owning office before escrow goes live on Day 31): buyer interviews confirm willingness to pay in writing from at least 10 design partners; escrow passes its testnet drill (refund, slash, pause); data procedures published; dispute paths staffed and drilled; metrics instrumented (NCV, verification cost, dispute rate, CAC by channel, human share, capital-to-move); Seer market templates configured and the first advisory markets drafted. Any unchecked item on Day 30 delays launch by one week per item.

Appendix D: arithmetic audit trail

Everything is computed by `model.py` from the inputs in its first block. Spot checks a reader can redo in a minute: contribution a task at AOV $60 = 0.109 × $60 − $0.95 = $5.59; break-even = $11,500 ÷ $5.59 = 2,057 tasks a month; Year-1 GMV = Σ tasks × AOV = $2,086,900; LMSR worst case = $850 × ln 2 = $589.18; capital to move a branch 3% at b = $850 = $850 × ln(0.5 ÷ 0.515) = $25.89; LTV/CAC = $111.51 ÷ $18.35 = 6.1×. Rounding: half-up; totals from unrounded values.

Appendix E: what changed from version 1 (6 November) and why

CHANGE	WHY
Upwork take-rate comparison corrected	Version 1 paired Marketplace revenue with total GSV (checked against the 10-K)
Bear and bull scenario figures replaced	Version 1's figures could not be reproduced from its stated inputs; `model.py` now derives them
§5.2 measured depth, §5.3 depth ladder, random slice, worthiness test, parent-normalised prices	Seer depth measured on 7 Nov (UTC) is far below version 1's trader and quorum assumptions; ideas credited to entries 9442, 9414, 9461/9424 and 9436
Recycled-worker netting; human-vs-agent split	Credited to entries 9477 and 9461
Lean case	The field's volume assumptions are about 1.7 to 7 times lower than version 1's ramp
§11 rewritten for 48 entries; "no other plan does this" statements removed	Version 1 compared 15 to 21 entries and several statements are no longer true
Mixed-language labels and table headings removed; sections renumbered	Editorial
Browser version with calculators; runnable demo and model in the entry kit	Verifiability
Decision rule uses a signed margin with a $5,000 floor; depth-ladder cap defined (stage cap, loss-based bound reported); Day-90 gate moved to 1,050 tasks and first binding decision *adopted*; Month-3 kill signal moved to Day 120; steward powers split into emergency and procedural; decision classes reconciled; T2 share defined as 10% of tasks; LTV/CAC attribution made explicit	Fixes from an independent review of the first version-2 draft
Launch float restated as a $104,700 raise (Days 1–30 and the audit are now funded explicitly); $20,000 standby removed	Version 1 omitted pre-launch costs and never defined the standby
Demo scenario, numbers and fixtures replaced; credits for entries 9363, 9374, 9385 and 9378 added (§11)	Several guard-rails and the demo pattern were first published by those entries

Independent proposal, prepared with AI assistance; disclosure is in this text and in the claim description, not in any image. No claim is made that Foreguild exists or has users, revenue or a treasury. Every figure is a cited fact, a labelled assumption, an estimate from `model.py`, or a measurement whose script is included.

PDF · model.py · tests · demo · depth data · depth script

Entry kit checks: `python3 -I model.py --check` · `python3 -I demo/run_demo.py` · `node test_js_parity.mjs`