



Build for the real world. Test in a virtual one.

Find implementation failures
before a healthcare pilot.

A three-month test of whether synthetic
workflow experiments help builders improve a design.

team@alivehealthtech.com





The same technology. Different workflows. Different outcomes.

Video visits were popular when the clinical and practical conditions fit. Yet the same study found sharply different outcomes across four services.

YOUNG ADULT DIABETES + CANCER SURGERY

About 20% of follow-ups moved to video.

Remote consultations became part of routine service delivery.

PRENATAL DIABETES + COMMUNITY HEART FAILURE

Abandoned in one.
Put on hold in the other.

Multidisciplinary handoffs and missing records strained one workflow; examination needs constrained the other.

A working interaction is not a working care pathway.

Before a live pilot, builders need to ask how their product behaves when staffing, handoffs, data and patient participation interact.



An empty alert queue can hide a broken product.

A monitoring startup tests the classifier together with device gaps, service latency and nurse shifts. 200 synthetic wearers, seven days and 30 paired runs expose a failure hidden by a quiet queue.

monitoring / experiment workspace

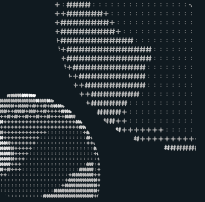
Local synthetic run

EXPERIMENT

world.json

models.lock.json

results.zip



Four variants.
Same world.
Matched seeds.

world.json / configuration excerpt

```
// Existing mock model + operating constraints
adapter_timeout_seconds: 5
latency_seconds: 20
fallback_to_threshold: false

staff: 1
shift_hours: 8
workdays: 5
```

TERMINAL

```
> alive run --world world.json --output results.zip
SUCCEEDED · 4 variants × 30 replications
Open metrics, assumptions and event traces.
```

Extracted results / modeled outcomes

Ideal conditions

98.8% detection

Among received abnormal samples; 50.3% of alerts are false alarms.

Latency stress

0% alerting

Replies arrive after the deadline. No alerts reach triage; its empty queue masks the failure.

Fallback retest

5.6 hours to review

Against an assumed 2-hour target. Mean queue wait for alerts starting review; unfinished waits excluded.



Put the workflow around the product.

Compose synthetic people, processes, staff capacity and service behavior around a proposed change. Run matched variants, inspect the event trail, change the design and retest.

Plan the right test.

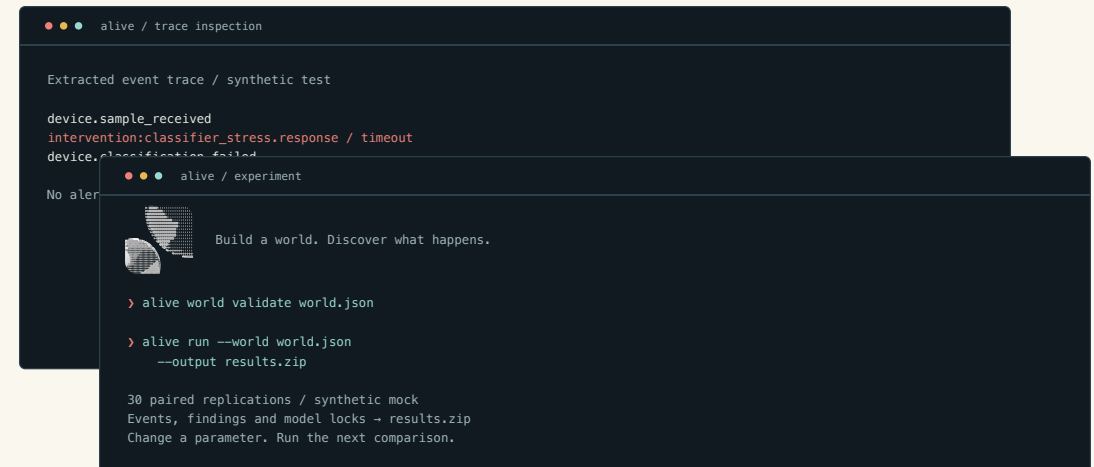
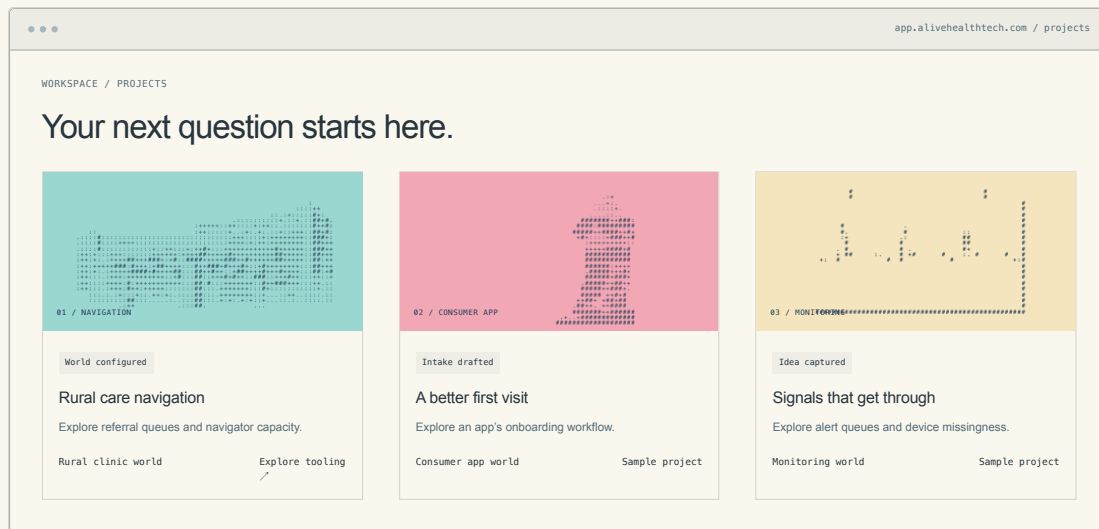
Specify the workflow and decision.

Build its environment.

Set demand, capacity and failure conditions.

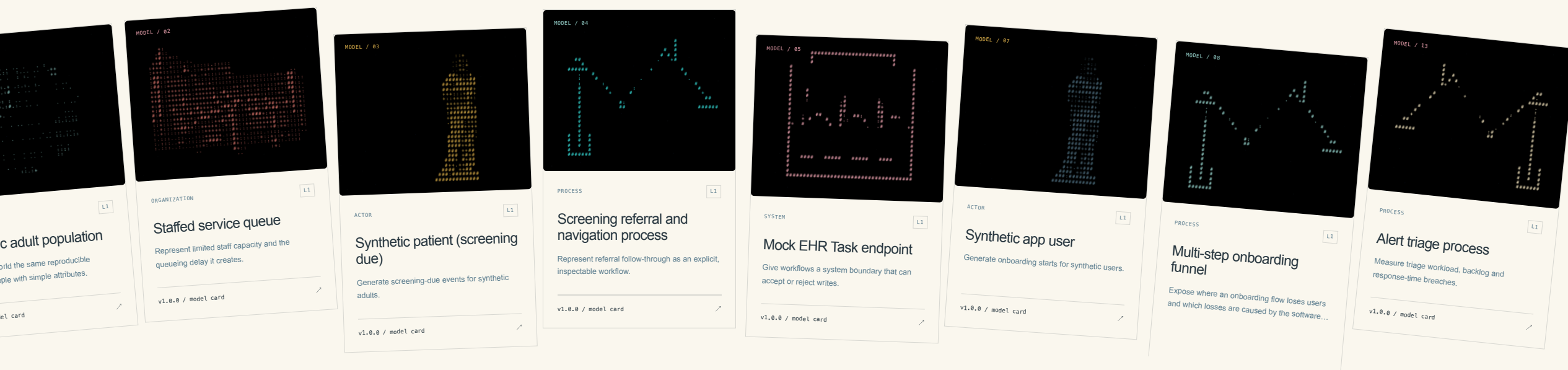
Inspect. Change. Retest.

Compare variants and preserve the trace.





Reusable models. Evidence you can inspect.



L1 • Executed

Observed behavior, event trace and assertion. The cards shown start here.

L2 • Internally verified

Replay, conservation, boundary and distribution tests pass for the stated use.

L3 • Calibrated

Fit to an external dataset; publish comparators, tolerances and applicability.

L4 • Externally validated

Independent back-test or prospective comparator; disclose predictive error.

Advance only with a versioned benchmark report and reviewer sign-off. Repeated use creates test cases; it does not establish accuracy.



Care is changing settings. The workflow has to keep up.

More tools in clinical work

72%

Report using at least one AI application.

AMA survey, 2026: 1,342 responses to the use-case question, up from 48% in 2024. Self-reported use; survey wording evolved.

More care outside hospital walls

Hospital-level care. Home-level constraints.

CMS's hospital-at-home study describes selection based on patients' circumstances, home environments and local resources. Technology must fit those conditions.

Evidence is an adoption requirement

88%

Say safety and efficacy validation matters.

AMA, 2026. Workflow experiments can inform what to test; external validation and live evaluation remain separate evidence gates.

The immediate opportunity: test the interactions around a new care workflow while its design can still change.



Bring implementation risk into the development loop.

Repeatable development-loop test ↑

Software behavior Antithesis Faults, execution and reproducible software tests. Synthea Repeatable synthetic patient-record generation.	Alive's intended opening Product + operating environment Reusable workflow models, technical failure paths and human/resource constraints in a run → inspect → retest loop.
Study inputs and assurance Synthea supplies records to studies. ORCHA assesses digital-health products.	Systems and behavior studies AnyLogic · Simul8 · FlexSim Operational and capacity simulation. Synthetic Users LLM-persona user research.

Technical faults / data

Operational / human behavior →

One-off study ↓



Start with one team preparing a monitoring pilot.

USER #1

A digital-health builder designing remote monitoring.

A founder or research engineer with an alerting prototype, a defined staff workflow and a near-term pilot. Their question: will a service failure or shift boundary silently break escalation?

Begin with synthetic inputs and the local CLI. Capture the decision, the experiment and the resulting design change.

3 teams

Proposed three-month cohort

Two early-stage digital-health product teams.

One university research team developing a monitoring workflow.

One workflow owner per team for assumptions and result review.

Recruit through labs and accelerator cohorts. Interview at least five builders before onboarding; record needs, consent and baseline decisions.



Built at the intersection of research and delivery.



Isaac Wilkins

Founder & President · Alive

Founder-led pilot: scenario design, engineering, onboarding and evaluation.

Harvard Medical School & Wheatley Institute

Co-investigator & Lead Researcher. Building a 250-intervention dataset on digital-health adoption and sustainment; translating implementation questions into testable assumptions.

MGH Healthcare Transformation Lab

Research Fellow. Led development and launch of two digital tools; AHA software collaboration and hospital-at-home technology-priority work.

Wildfire Partners VC

Founding Team; Advisory Board. Fund formation, investor materials and digital-health investment strategy.

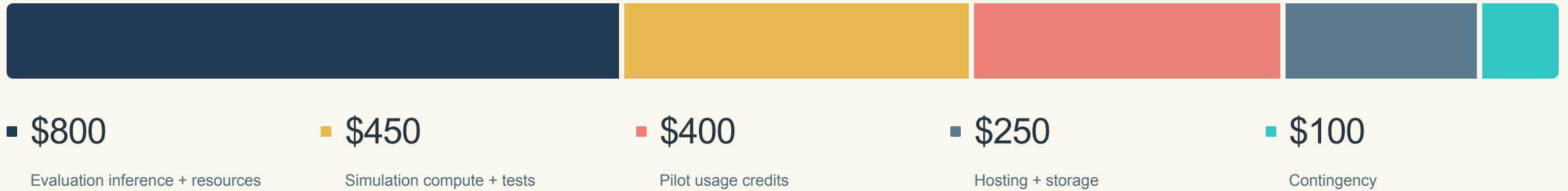
McKinsey & Company

Incoming 2027 Summer Business Analyst. Future role, after this proposed three-month pilot.

Scope the pilot to three teams and one workflow. The \$2,000 funds operating expenses, with \$0 personnel; recruiting roles are outside the funded delivery plan. Weekly availability and contributor arrangements must be confirmed before launch.



\$2,000 to test whether builders come back.



Month 1 · Establish the benchmark

Replay + explain.

Publish a locked monitoring benchmark, boundary tests and parameter provenance; interview ≥5 builders.

Month 2 · Make it usable

Onboard 3 teams.

Each runs a baseline and changed variant, inspects a trace and records a design decision.

Month 3 · Test the value

Change. Return. Rerun.

≥1 team changes a design after a finding; ≥2 teams return to test a later version.

Deliver the evidence, even if the pilot misses its targets.

Report runs, decisions, return use, actual spend and barriers. This grant tests usefulness; external calibration is a separate workstream.



Recurring access. Revenue that expands with usage.

Developer

\$39 /mo

\$10 included cloud credits

Team workspace

\$149 /mo

\$40 included cloud credits

Institutional pilot

\$500 /mo assumed

Custom packaging to test

Bottom-up subscription revenue scenarios

Adoption scenario	Dev / Team / Inst.	MRR	Annual run rate
Initial paid cohort	100 / 20 / 2	\$7,880	\$94,560
Repeatable adoption	500 / 100 / 10	\$39,400	\$472,800
Expanded adoption	1500 / 300 / 30	\$118,200	\$1,418,400

Beyond subscriptions

More tests.
More consumption.

Meter usage beyond included credits. Institutions can sponsor cohorts or commit capacity.

~60% run contribution

At the draft's assumed 2.5× direct run-cost billing rate, before account costs and overhead.

Counts are hypothetical paid accounts / workspaces / pilots, with no acquisition timeline. Excludes overages, discounts and churn; institutional delivery cost is not modeled.



Model the margin. Stress-test the assumptions.

Monthly unit economics · specified usage profiles

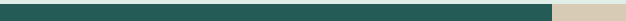
Account	Revenue	Run + account cost	Contribution
Active free	\$0	\$0.59	-\$0.59
Developer	\$39	\$4.62	\$34.38 / 88.1%
Team workspace	\$149	\$17.79	\$131.21 / 88.1%

Direct cost / run \$0.02 Light \$0.20 Standard \$2.11 Advanced

Paid subscription contribution

88.1%

500 Developers + 100 Teams
\$34,400 monthly revenue
\$4,090 modeled costs



Institutional costs excluded

Free-tier subsidy

86.4% contribution

Same paid mix + 1,000 active free accounts: \$589/mo subsidy.

Usage stress case

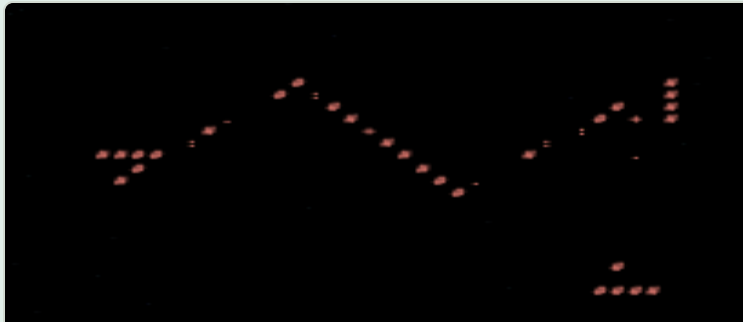
78.3% contribution

Same runs; 20% worker utilization + 2× inference price; free subsidy included.

Base case: 40% worker utilization. Free / Dev / Team run mix: 6/1/0 · 15/7/1 · 60/30/3 (light/standard/advanced). Contribution must fund the unmodeled operating budget.

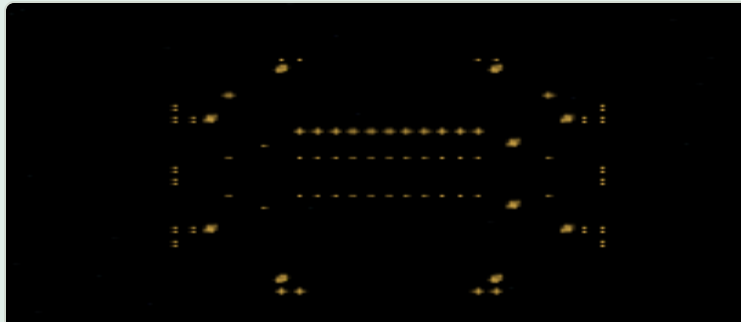


New industries. The same testing loop.



Manufacturing

Where will the production line back up?



Infrastructure

What happens when a shared resource fails?



Consumer developers

Where will users drop out of my app?

Start with healthcare.
Build toward the world around any product.

Domain-specific models.
A familiar way to ask, test and iterate.



Start free. Run a useful test. Come back.

Try

Free local
testing.

CLI, docs, open methods.

Return

Test every
new version.

Make it a habit.

Expand

Bring your
team.

Shared experiments.

Sponsor

Fund a whole
cohort.

Universities + accelerators.



What is built. What the pilot adds.

Built and exercised locally

CLI world validation, four-variant experiments, seeded trials and ZIP exports.

Versioned reference models, metrics, assertions and event traces.

The monitoring stress experiment: recorded outputs and exact replay tests.

Portal workflow code exists; deck dashboard and IDE compositions are previews.

Next, with explicit gates

Onboard three builders using one defined monitoring workflow.

Record design changes, return sessions and actual operating costs.

Calibrated external benchmarks, live staging/FHIR and IDE integrations remain future work.

Confirm discovery participants, founder schedule and any contributor terms before launch.

The demo establishes modeled behavior under synthetic assumptions. It does not establish a real classifier's accuracy, clinical benefit or a calibrated hospital twin.



Test the mechanics. Then test the match to reality.

1. Verify the mechanism.

Freeze versions and seeds. Check replay, event order, conserved counts, zero capacity and extreme demand. Publish failures alongside passing assertions.

2. Calibrate a stated use.

Choose published operational data with appropriate rights. Predeclare outputs and tolerances, fit parameters, retain a holdout and document where the model applies.

3. Challenge the prediction.

Back-test a known implementation outcome with an independent comparator. Report error, sensitivity and mismatches; obtain reviewer sign-off before promoting evidence.

The three-month grant tests usefulness to builders. Calibration and predictive validation require their own comparator data, review and resources.



A reproducible monitoring stress test.

Common world across four variants

Input	Specified scenario	Scope
Population / run	200 synthetic wearers · 7 days · 30 paired replications · seed 42	Unvalidated synthetic population
Device behavior	4 samples/day · 10% non-wear · 10% transmission loss	Independent assumed gaps
Synthetic abnormal label	3% of generated samples	Not a physiological model
Staff / response target	1 staff member · 8-hour shifts · 5 days/week · 2-hour target	Fixed across variants
Candidate classifier	99% assumed sensitivity · 3% false-positive probability	Mock service; not a real classifier
Stress / fallback retest	20-second replies vs 5-second timeout · 15% service errors	Fallback enabled only in retest

Ideal-condition candidate: 0.2-second latency, no errors. Retest also lowers assumed false-positive probability to 0.5%; with every reply timing out, that classifier setting has no effect in this case. The baseline is an idealized perfect threshold on received abnormal samples.

Reproduce: `uv run alive run --world research/demos/monitoring-stress/world.json --output monitoring.zip`

Evidence record: `research/demos/monitoring-stress/summary.json · 30 replications per variant · run run_cfef03c7a38829b7`



The failure path changes what the dashboard means.

Selected variant metrics

Metric	Ideal candidate	Latency stress	Fallback retest
Received abnormal samples alerted	98.8% [98.5%, 99.1%]	0.0% [0.0%, 0.0%]	100.0% [100.0%, 100.0%]
False alerts / all alerts	50.3% [49.1%, 51.5%]	Undefined / no alerts	0.0% [0.0%, 0.0%]
Generated abnormal samples alerted	89.1% [88.3%, 89.9%]	0.0% [0.0%, 0.0%]	90.2% [89.4%, 91.0%]
Queue wait / started alerts, minutes	372.5 [365.2, 379.9]	Undefined / no alerts	338.8 [329.1, 348.4]
Response-target breaches / day	30.6 [29.8, 31.4]	0.0 [0.0, 0.0]	14.3 [13.8, 14.9]

Zero response-target breaches during the latency stress case means no alerts reached triage. It does not mean the workflow improved.

Both structural assertions pass in every variant: alerts have source events; adapter requests settle exactly once. These assertions do not require successful alerting. Queue waits exclude unfinished waits; non-wear is outside the detection denominator. Intervals quantify replication noise, not parameter or clinical uncertainty.



From workload assumptions to dollars per run.

Illustrative run profiles · USD

Input	Light	Standard	Advanced
Worker / productive time	1 vCPU, 2 GiB / 4 min	1 vCPU, 2 GiB / 30 min	4 vCPU, 8 GiB / 180 min
Worker at 40% utilization	\$0.00298	\$0.02233	\$0.53573
Luna input / output tokens	10k / 2k	100k / 20k	500k / 100k
Terra input / output tokens	0 / 0	20k / 4k	250k / 50k
Combined inference	\$0.00440	\$0.13200	\$1.32000
Artifact / network allowance	\$0.01500	\$0.05000	\$0.25000
Total direct cost	\$0.02238	\$0.20433	\$2.10573

`worker hours × listed hourly rate ÷ utilization + inference + artifact allowance`

Worker rates: \$0.01786/hour (1 vCPU/2 GiB), \$0.07143/hour (4 vCPU/8 GiB). Inference per million input/output tokens: Luna \$0.20/\$1.20; Terra \$2/\$12. Token totals span requests of at most 272k input tokens; long-context premiums and cache pricing are excluded. GPU, web/tool fees, licenses and other overhead are excluded.



\$2,000 in cash. A deliberately bounded pilot.

Proposed monthly cash allocation

Expense	Month 1	Month 2	Month 3	Total
Evaluation inference / permitted resources	\$300	\$300	\$200	\$800
Simulation compute / engineering tests	\$100	\$150	\$200	\$450
Subsidized pilot usage	\$0	\$100	\$300	\$400
Hosting / database / storage	\$75	\$75	\$100	\$250
Contingency	\$0	\$0	\$100	\$100
Total	\$475	\$625	\$900	\$2,000

What the pilot should demonstrate

Three teams complete paired experiments; at least one changes its design; at least two rerun a later version. Publish a pilot report with actual costs and missed targets.

What the budget does not buy

Founder salary, foundation-model training, a large free-cloud cohort, clinical validation, regulatory clearance or a calibrated cross-industry twin.



Evidence with its denominator and boundary.

Problem and timing

Greenhalgh et al., JMIR 2017

Four video-consultation services; differing adoption outcomes and workflow constraints.

AMA Physician AI Survey, 2026

72% explicit use, 1,342 responses to the use-case question; 88% rate validation important. Self-report; changing survey instrument.

CMS hospital-at-home study, 2024

Local resources and home context affect program selection.

Product and pilot evidence

Committed demo world, model lock, recorded summary and reproducibility notes: [research/demos/monitoring-stress/](#).

Evidence definitions: canonical product spec, §2. Local model cards remain L1; no external calibration claimed.

Financial assumptions

Appendix retains proposed pricing and hypothetical adoption scenarios from the original deck. Decimal precision is arithmetic, not measured costs or traction.

Provider rates: historical planning inputs checked October 13, 2026. The complete notes include [DigitalOcean](#) and [OpenAI](#) source links.



Unit-economics assumptions and reproducible arithmetic

Important: These are transparent **planning scenarios** for a future product, not Alive usage logs. Input/provider list prices checked October 13, 2026. They intentionally exclude speculative market-adoption claims. Recheck before external publication.



A1. Source price assumptions

Worker CPU/memory: DigitalOcean Basic Droplets: **1 vCPU / 2 GiB = \$0.01786/hour**; **4 vCPU / 8 GiB = \$0.07143/hour**. [P1]

Effective productive utilization: **40%** assumed for short-lived workers, with the remaining paid capacity allocated to idle time, startup, scheduling, and imperfect packing. Effective CPU-worker-hour rate = listed rate / 0.40. An actual managed job platform may bill differently.

Budget LLM: OpenAI GPT-5.6 Luna: **\$0.20 / million input tokens**; **\$1.20 / million output tokens**. [P2]

More capable LLM: OpenAI GPT-5.6 Terra: **\$2.00 / million input tokens**; **\$12.00 / million output tokens**. [P3]

Token batching assumption: The A2 token volumes are totals across calls. Each request stays at or below the provider's 272k-input-token long-context threshold; a single 500k-token Luna prompt would attract higher rates and is not priced by this base case. Cache write/read pricing is not modeled. [P2]

Per-run storage/network/orchestration allowances: **\$0.015 lightweight**, **\$0.05 standard**, **\$0.25 advanced**, *assumptions*, not provider invoices. Includes reasonable buffer for artifacts/network incidental usage, but not all possible vendor/tool charges.



A1. Source price assumptions / continued

Persistent object storage reference: DigitalOcean Spaces starts at **\$5/month with 250 GiB included**. [P4] Persistent baseline services should be tracked separately and not charged twice in run-level costing.

No web-search/tool fees, dedicated GPU, third-party license fees, managed database minimums, payment processing, 24/7 dedicated workers, enterprise support labor, or compliance/legal expenses are included in **per-run** calculations. Unusually memory-intensive tasks would require a larger worker tier.



A2. Direct cost per illustrative run

Cost item	Lightweight	Standard	Advanced
Worker specification	1 vCPU/2 GiB	1 vCPU/2 GiB	4 vCPU/8 GiB
Worker time	4 min	30 min	180 min
Worker cost at 40% utilization	\$0.00298	\$0.02233	\$0.53573
Luna token use (input/output)	10k / 2k	100k / 20k	500k / 100k
Luna inference	\$0.00440	\$0.04400	\$0.22000
Terra token use (input/output)	0 / 0	20k / 4k	250k / 50k
Terra inference	\$0.00000	\$0.08800	\$1.10000
Artifact/network/orchestration allowance	\$0.01500	\$0.05000	\$0.25000
Total direct cost	\$0.02238	\$0.20433	\$2.10573
Slide-rounded price	\$0.02	\$0.20	\$2.11



A2. Direct cost per illustrative run / continued

Formula: $(\text{worker_hours} \times \text{listed_worker_hourly_rate} / 0.40) + \Sigma[(\text{input_tokens}/1\text{M} \times \text{input_rate}) + (\text{output_tokens}/1\text{M} \times \text{output_rate})] + \text{artifact_allowance}$.

Interpretation: A standard numerical simulation without LLM calls could cost much less; a large agent-based evaluation with hundreds of thousands of expensive model calls could cost far more. These illustrative configurations are deliberately explicit, not universal averages.



A3. Three monthly user/workspace profiles

Scenario	Lightweight	Standard	Advanced	Modeled run expense	Allocated fixed account expense ¹	Cost / month	Subscription
Active free account	6	1	0	\$0.34	\$0.25	\$0.59	\$0
Paid Developer	15	7	1	\$3.87	\$0.75	\$4.62	\$39
Paid Team workspace	60	30	3	\$13.79	\$4.00	\$17.79	\$149

¹ **Fixed account expense** is an assumed allocation toward persistent application hosting, database, observability, and storage—not an invoice or measured marginal cost. Team is one **workspace**, not one seat; Dev is one account. Costs and dollars are rounded only at display stage.

Proposed credits at **customer-facing billing value: Community \$2, Developer \$10, Team \$40** each month. These are **not** the amount Alive necessarily spends. Unused credits expire in this illustrative setup. Paid plans have prepaid overages or pay-as-you-go limits.



A3. Three monthly user/workspace profiles / continued

For analysis, suppose usage is valued at **2.5× underlying modeled run expense** (subject to minimum charges, premium models, taxes, and future actual cost calibration): Community's sample consumes **about \$0.85** retail; Developer **about \$9.68**; Team **about \$34.47**, each within the stated allowance. This markup is a price **hypothesis**, not a proven market-clearing rate.

The Developer profile costs **\$4.62 / \$39** (about **88.1%** modeled *pre-personnel direct contribution margin*); the Team profile costs **\$17.79 / \$149** (about **88.1%**). These are not financial-statement gross margins, and don't incorporate normal enterprise customer-success/engineering overhead.

The Community monthly cost applies to an **active free account with the specified usage**. Inactive accounts have different costs; heavy users must be capped, BYOK, or moved to paid/prepaid credits. If all 1,000 free pilot users behaved like this active profile, that alone would imply **about \$590/month** variable and allocated costs; our **\$400 three-month grant subsidy** cannot fund that volume. Therefore the pilot must be smaller, local-first, BYOK where possible, and budget-capped.

If hosted runs are billed at a resource rate below actual utilization-adjusted and support expense, Alive loses money on consumption. Enforce run-level **preflight cost estimates, concurrency caps, reservations/refunds for credits, hard spend limits, metering, and alerts**.



A4. Break-even intuition (illustrative)

At **\$39/month** with an illustrative **\$4.62/month** account-level run + allocated expense, the remaining **\$34.38** can help cover shared fixed costs, customer acquisition, and other unmodeled expenses. This does not mean the company breaks even on one customer. A hypothetical **\$200/month** in incremental shared fixed costs would require **6 similarly active Developer accounts** to cover that amount *under the scenario*, assuming no additional unmodeled costs or team subscriptions. True break-even requires full P&L, not just marginal run costs.



A5. Suggested pilot metrics to replace assumptions

Collect per-run worker-time, actual billed inference input/output, retry rates, memory profile, generated artifact bytes, cloud idle capacity, customer-tier active usage, dollar cost by scenario, and model output/replay quality. Publish internal dashboards and recompute economics monthly.



A6. Portfolio contribution and sensitivity (October 13 revision)

All costs below use A1–A3 without rounding intermediate calculations. Implementation: `apps/site/app/deck/economics.ts`; tests reconcile the source costs and revenue case.

Scenario	Monthly revenue	Modeled monthly expense	Contribution	Contribution rate
500 Developers + 100 Team workspaces	\$34,400	\$4,089.78	\$30,310.22	88.1%
Same paid mix + 1,000 active free accounts	\$34,400	\$4,678.36	\$29,721.64	86.4%
Same accounts/runs; 20% productive utilization and 2× inference prices	\$34,400	\$7,464.23	\$26,935.78	78.3%

The free subsidy is \$588.59/month for 1,000 active free accounts at the stated profile. The stress case doubles utilization-adjusted worker expense and inference expense; artifact allowances and fixed account allocations remain constant. It does not double run volume. Institutional revenue and delivery cost are excluded from this contribution calculation; the separate \$39,400 MRR example includes 10 institutional pilots at an assumed \$500/month. These are pre-personnel direct contribution estimates, not gross margins or company profitability. Payroll, support, development, acquisition, payment fees, legal/security and other overhead must be funded from contribution. Customer counts, willingness to pay, conversion, retention, and institutional costs are unvalidated; there is no dated operating forecast or LTV/CAC claim.



Three-month \$2,000 grant cash plan

Allocation	Month 1	Month 2	Month 3	Total
Evaluation inference / permitted resources	\$300	\$300	\$200	\$800
Simulation compute / load testing	\$100	\$150	\$200	\$450
Capped pilot execution credits	\$0	\$100	\$300	\$400
Hosting / databases / storage	\$75	\$75	\$100	\$250
Contingency	\$0	\$0	\$100	\$100
Total	\$475	\$625	\$900	\$2,000

Milestones

Month 1 — Replay and explain: Publish a version-locked monitoring benchmark, boundary tests and provenance. Complete at least five builder interviews.



Three-month \$2,000 grant cash plan / continued

Month 2 — Onboard three teams: Each completes a baseline and changed variant, trace review and documented design decision.

Month 3 — Change and return: At least one team changes its design because of a finding; at least two return in a separate session to test a later version. Publish actual spend, outcomes and barriers.

Boundaries: No talent/payroll allocation; no assumption of 1,000+ funded cloud users; no clinical PHI collection by default; no claim of real-world clinical predictions; support only allowable grant expenditures. Reallocate within approved budget only if the sponsor allows it.



Executed synthetic demo and exported appendix

The core export contains ten grant-focused slides. The full export adds eleven seafoam appendix slides, then the complete detailed assumptions, arithmetic, grant cash plan and linked sources. The optional equity-fundraising template is excluded.

Demo reproduction and interpretation: `research/demos/monitoring-stress/README.md`. WorldSpec: `research/demos/monitoring-stress/world.json`. Evidence record: `research/demos/monitoring-stress/summary.json`; exact resolved models: `research/demos/monitoring-stress/models.lock.json`. Replication-zero event digests are replayed in automated tests. No production model, default, contract or canonical fixture changes.

Metric	Ideal candidate	Latency stress	Fallback retest
Received abnormal samples alerted	98.8% [98.5%, 99.1%]	0% [0%, 0%]	100% [100%, 100%]
False alerts / all alerts	50.3% [49.1%, 51.5%]	Undefined: no alerts	0% [0%, 0%]
Generated abnormal samples alerted	89.1% [88.3%, 89.9%]	0% [0%, 0%]	90.2% [89.4%, 91.0%]
Mean wait among started alerts, minutes	372.5 [365.2, 379.9]	Undefined: no alerts	338.8 [329.1, 348.4]
Response-target breaches / simulated day	30.6 [29.8, 31.4]	0 [0, 0]	14.3 [13.8, 14.9]



Executed synthetic demo and exported appendix / continued

All quantities are **modeled_outcome** from an executed mock experiment, not clinical validation. Intervals represent replication variation, not uncertainty in parameter assumptions. See the README for non-wear exclusions, censored waits, idealized comparator, fixed latency, variant changes and next tests that have not been executed.



Implementation evidence

Greenhalgh et al., *Beyond Adoption*, JMIR 2017, Case A: four clinical services. Young adult diabetes and cancer surgery used video for about 20% of follow-ups; prenatal diabetes abandoned it and community heart failure paused it. This is a qualitative case study, not a population-wide failure rate, and is not an Alive prediction or the founder's dataset result. [Original study.](#)

The founder's 250-intervention adoption/sustainment dataset is in development. No aggregate finding from that dataset is claimed in this deck.



Healthcare timing

AMA 2026 Physician AI Survey, March 2026, pages 4–5: 72% report incorporating at least one AI use case, versus 48% in 2024. The use-case question has 1,342 responses in 2026; 1,183 in 2024. Do not substitute 81% awareness-or-use for explicit use. Total survey participants: 1,692; denominators vary by question. Survey changes limit strict comparisons across years. Page 4 reports 88% regard safety/efficacy validation as important to adoption. These are physician self-reports, not verified deployment or Alive demand.

CMS hospital-at-home study fact sheet, September 30, 2024: program selection depends on clinical and psychosocial circumstances, home/community environment and local resource capabilities. This historical study supports the workflow argument; the deck does not claim current waiver eligibility or a regulatory requirement for Alive.



Competition map

The map is editorial positioning, not measured coordinates or a claim that adjacent tools cannot support repeated tests. Simulation packages can automate experiments. Synthea spans repeated generation and study inputs, so appears in both left quadrants. The right-upper cell describes Alive's intended product workflow; the status slide distinguishes implemented local experiments from planned integrations.

Antithesis: reproducible software simulation testing.

Synthea: open-source synthetic patient and record generation; a complementary fixture source, not an operational workflow twin.

AnyLogic: multi-method simulation.

Simul8: healthcare process simulation.

FlexSim Healthcare: operational simulation and capacity studies.

Synthetic Users: AI persona research.

ORCHA: digital-health assurance.



Modeled cost source inputs

Historical list-price inputs, checked October 13, 2026. Recheck before relying on them; calculation precision is not measurement accuracy.

[P1] DigitalOcean Basic Droplets.

[P2] OpenAI GPT-5.6 Luna.

[P3] OpenAI GPT-5.6 Terra.

[P4] DigitalOcean Spaces.



Pilot decisions and scope

Recruitment and success counts are proposed targets, not signed partners or traction. Five discovery interviews precede three participating teams. Completion means a baseline and changed variant, a recorded trace review and a design decision. A changed design must link to an experiment finding. A return means a separate session testing a later version by month three, not another click during onboarding.

Budget: \$800 for evaluation inference and permitted evaluation resources; \$450 simulation compute and tests; \$400 capped pilot usage credits; \$250 hosting/database/storage; \$100 contingency. Resource purchases depend on sponsor eligibility and demonstrated need. No contractor or personnel compensation is budgeted. Record receipts and metered consumption; unused credits are not automatically spend. Stop onboarding or cloud runs before exceeding the cap.

Founder weekly availability, recruitment-role compensation and any additional contributors remain unconfirmed. Neither interviewing candidates nor past institutional work is counted as pilot staffing. The pilot must have a founder-owned schedule before launch; optional hires are outside its critical path.



Evidence progression

L1–L4 are internal terms scoped to a metric, population and use. L1: execution with traces and assertions. L2: internal replay, conservation, boundary and distribution tests. L3: calibration to documented external data with tolerances and applicability. L4: independent retrospective or prospective validation with disclosed predictive errors. Evidence promotion requires a versioned report and reviewer sign-off; no model grade is promoted by this deck.

Choose a published operational comparator with usable data and rights, declare outputs and tolerances before fitting, retain a holdout, report discrepancies and sensitive assumptions, then obtain independent review. A back-test against a known implementation outcome is useful only when the modeled mechanisms and population fit. The grant does not promise clinical validation or patient benefit.