External Test Lab & Community DevNet

4-month pilot proposal for community-owned testing infrastructure and QA capacity

1. Executive Summary

This proposal requests funding for a 4-month pilot of External Test Lab & Community DevNet: a community-owned testing function for Gonka protocol upgrades, DevShards, inference flows, host/broker operations, and geographically distributed network behavior before governance decisions and production rollout.

Item	Proposal
Pilot duration	4 months
Requested budget	Up to 22,000 USDT per month, plus a one-time 80,000 GNK end-of-pilot recognition payment.
Funding model	Monthly tranches: Month 1 prepaid, Months 2–4 released after previous-month deliverables are accepted
Main deliverables	Community DevNet, external testing team, public reports, test plans, runbooks and issue tracker
Ownership	Community-owned infrastructure and open operational artifacts, with clear handoff plan
Security handling	Public-by-default reporting, with private disclosure for security-sensitive findings before remediation

2. Problem Statement

Today, Gonka lacks a dedicated community-owned validation layer for important network changes. The External Test Lab and Community DevNet are proposed to close this gap by providing practical testing capacity, shared infrastructure, and public evidence before releases, governance decisions, or production rollout.

The External Test Lab and Community DevNet will:

- Validate protocol upgrades, DevShard releases, broker/inference flows, integrations, and other critical network changes before they move forward.
- Test distributed-network behavior that is hard to verify in local or internal environments, including latency, synchronization, propagation, and regional instability.
- Use available testing capacity for proactive bug hunting, regression checks, and investigation of known weak points when no release candidate is waiting for validation.
- Provide a neutral testing path for work delivered by external teams and ecosystem contributors before it is accepted, funded further, or used in production.
- Help validate vulnerability reports from researchers, audits, or programs such as HackerOne through reproduction, impact assessment, regression testing, and confirmation after remediation.
- Provide a shared environment where trusted hosts and teams can safely test proposals, integrations, DevShard scenarios, protocol behavior, and early implementation ideas.
- Produce clearer testing evidence for governance participants, including test plans, dashboards, runbooks, issue trackers, defect reports, security-sensitive disclosure handling, and release-readiness summaries.

This adds a missing validation layer for the Gonka ecosystem while complementing Core Team testing.

3. Roadmap Alignment

This proposal directly implements two projects from the Gonka Network Development Roadmap.

Track 4. Network reliability and observability — Project 2. External testing lab
The roadmap defines an external testing lab for Gonka changes before broad rollout, including changes from Protocol Maintainers, funded external teams, and ecosystem contributors.

This proposal implements that project through the External Testing Team, test plans, smoke and regression checks, defect reports, public issue tracking, and release-readiness reports.

Track 7. Public sandbox and consumer-GPU testnet — Project 1. Public testing sandbox
The roadmap defines a separate test environment for experiments with models, parameters, integrations, DevShard scenarios, protocol-level behavior, validation, settlement, and upgrade testing before mainnet.

This proposal implements that project through Community DevNet: a small, always-on, geographically distributed network for protocol, node, DevShard, operational, integration, and distributed-behavior testing.

4. What We Are Building

The project has three connected components.

Component	Purpose	Main output
Community DevNet	Small always-on geographically distributed network for protocol, node, DevShard, and operational testing.	9–13 inference nodes, monitoring dashboard, deployment runbook
Burst GPU Testing Budget	Temporary rental of large GPU infrastructure when heavy model or load testing is needed.	Monthly rental allowance, test logs, benchmarks, and cost report
External Testing Team	Two hands-on QA / infrastructure testing engineers to validate pre-release builds, DevShards, and deliverables from external teams.	Test plans, smoke/regression checks, defect reports, readiness reports

5. Community DevNet Infrastructure

- Target size: 9–13 always-on inference machines plus required network nodes, where feasible within the monthly DevNet infrastructure cap.
- Topology: part of the DevNet runs as Network Nodes with multiple attached MLNodes, so that realistic multi-MLNode host configurations can be reproduced and tested.
- Indicative distribution: North America East, North America West, United Kingdom, Germany, France, Finland, and Asian locations depending on network quality and hosting availability.
- Model profile: DevNet inference nodes are expected to run lightweight instruct models, such as Qwen/Qwen3-0.6B, or equivalent models.
- Inference nodes: NVIDIA GPU machines with 16 GB VRAM, compatible with the project's CUDA 13.0 container/runtime stack.
- Goal: protocol and distributed behavior testing.
- Monitoring: public dashboard for node availability.
- Operations: infrastructure lead from the host/DevOps community responsible for provisioning, monitoring, and maintenance.

Public access. The DevNet is intended to serve the broader community, not only the testing team. During the pilot, access for external participants is granted through a lightweight request process — primarily to manage abuse,

access control, and node stability while monitoring and onboarding documentation are still being built. Intended external use cases include:

- hosts rehearsing onboarding and node operations before joining mainnet;
- developers and users experimenting with inference, smart contracts, and integrations outside the QA team's test plan.

Broader participation is the target state: we intend to move to fully permissionless access as soon as safety, abuse limits, onboarding documentation, and monitoring allow it.

6. Burst GPU Testing Budget

- Used only when needed for release candidates, large model tests, load tests, and model compatibility checks.
- Planning assumption: Up to one calendar week (168 hours) of rental per month.
- For Kimi-class testing, the requirement estimate assumes an ML Node with either 4× NVIDIA B200 or 8× NVIDIA H200, around 640 GB total GPU VRAM, 960 GB+ system RAM, 16-core amd64 CPU, and a Network Node host with 16-core CPU, 64 GB+ RAM, 1 TB NVMe, and at least 100 Mbps networking.

Access and accountability. Burst GPU capacity is provisioned and coordinated by the External Testing Team together with the project owners, primarily for release-candidate validation, model compatibility, and load tests. Requests from Protocol Maintainers and ecosystem teams are accommodated where capacity allows. All burst usage is itemized in the monthly public report: purpose, hours used, and cost per test run, with test logs published alongside.

7. External Testing Team

The proposal funds two external QA / Infrastructure Testing Engineers.

Scope of work

- Contribute to quality standards and test strategy for the network
- Define acceptance criteria and test plans for core network lifecycle events and components
- Build pre-release validation frameworks, including smoke checks and regression coverage for known failure patterns
- Deploy and verify distributed node and service stacks in production-like environments
- Validate critical cross-system flows end to end, with documented evidence and clear defect escalation — e.g. participant and key flows, inference and proof-of-compute participation, gateway and proxy behavior
- Use health signals, chain queries, and logs as primary validation inputs, not just debugging aids
- Verify upgrade readiness, rollback feasibility, and post-change health across the stack
- Establish, tune, and document practical operational primitives, such as PoC validation threshold and inference validation threshold settings, based on DevNet experience, and contribute these findings directly to the official Gonka documentation where appropriate.
- Support DevNet validation and release-readiness assessment before production rollouts
- Validate recovery and incident resolution through root-cause analysis and re-testing
- Report defects with clear reproduction steps, impact assessment, and release-blocking status
- Track and communicate quality metrics that reflect network health and operational reliability

Required capabilities

- 3+ years in system QA / Test Engineering, SDET, or quality-focused DevOps/SRE
- Understanding of blockchain lifecycle, key and auth flows: registration, delegated permissions, fee grants

- Experience in operation and validation of blockchain nodes (Cosmos SDK preferred): sync, recovery, RPC queries, network phase behavior
- Experience validating distributed systems in production-like environments
- Strong test design: test plans, acceptance criteria, smoke/regression/e2e, edge cases, negative testing
- Solid Linux, SSH, shell scripting, and log-based verification
- Docker & container orchestration - including environment and config reload behavior
- Clear defect reporting and documentation — runbooks, test results, sign-off checklists
- Comfort with time-boxed validation before critical network events

Preferred / Nice-to-Have

- SDET experience: scripted validation, CI pipelines, automated health checks
- Blockchain / Web3 QA: testnet operations, bridge testing, wallet/key flows, upgrade regression
- Cross-chain testing: EVM testnet validation, withdrawal flows, contract interaction
- GPU/ML inference testing: worker health, model serving, artifact delivery
- Experience with coordinated multi-node upgrades
- API gateway and proxy testing (failures, latency, request tracing)
- Decentralized inference or Proof of Compute networks

8. Project Owners and Accountability

Role	Proposed owner	Responsibilities
Project Lead	Sergii Paranko (S∃ga L∈nin)	Overall project accountability, governance coordination, scope and milestone control, QA contractor selection and onboarding, public reporting, and testing laboratory oversight.
Infrastructure Lead	Mikhail Chudinov (Mitch)	DevNet provisioning, hardware selection, regional hosting, monitoring, operational stability, and infrastructure cost control.
External Testing Team	2 hired QA / Infrastructure Testing Engineers	Test execution, reports, defect documentation, regression tracking, and release-readiness recommendations.

9. Transparency and Reporting

- Public dashboard for DevNet health and node status.
- Public task board for planned, active, and completed testing work.
- Public issue tracker for non-sensitive bugs, regressions, and operational findings.
- Monthly public report with deliverables, incidents, spending by budget line, remaining balance, unused funds, and next-month plan.
- Per-release readiness report before governance vote or production rollout when a release candidate is provided in time.
- Security-sensitive findings are reported privately to Core Team first; a public placeholder issue is created where appropriate, and details are disclosed after remediation or agreed disclosure window.

10. Protocol Maintainer Coordination

- The External Test Lab is intended to work in coordination with Protocol Maintainers while remaining an external community testing function.

- Protocol Maintainers are expected to support initial onboarding by providing technical context, relevant documentation, general guidance, scripts and notes, expected test focus, and clarification of protocol-specific behavior where needed.
- This coordination helps the testing team become productive faster and reduces the risk of misinterpreting expected network behavior. At the same time, validation reports remain independently prepared by the External Test Lab and are published for the community.

11. Release Validation Handoff Requirements

Validation type	Expected handoff
Protocol release / production upgrade	Protocol Maintainers provide release candidate, upgrade notes, affected components, and expected test focus at least 7 days before the planned governance vote or production rollout, where feasible.
DevShard / test environment validation	DevShards or another testable environment are provided with scope and expected test focus at least 3 days before the expected validation result, where applicable.

Where a governance vote or production rollout is scheduled and testable artifacts are provided in time, the External Test Lab publishes a readiness report covering tested areas, pass/fail results, known risks, and recommendations.

If testable artifacts are provided late or incomplete, the External Test Lab may still perform limited validation, but the report will clearly state the reduced scope, time constraints, and known limitations.

12. Milestones and Acceptance Criteria

Workstream A: DevNet Infrastructure

Milestone	Timing	Milestone result	Public artifacts	Payment gate
M1: Setup	Month 1	DevNet design agreed, infrastructure procurement started, at least 5 nodes online; infrastructure blockers documented if any.	DevNet architecture note; initial status dashboard (live link); draft node deployment runbook.	Month 2 funding requires M1 report and acceptance.
M2: Full DevNet operational	Month 2	At least 9 nodes online across target regions; monitoring in place.	Published node deployment runbook sufficient to reproduce a DevNet node; public dashboard showing all nodes; regional layout summary.	Month 3 funding requires M2 report and acceptance.
M3: Stable DevNet operation	Month 3	DevNet running stably; operational issues identified and addressed.	Updated runbook; incident log for the month; onboarding guide for external DevNet participants, interim infrastructure cost report.	Month 4 funding requires M3 report and acceptance.
M4: Pilot completion and reporting	Month 4	Pilot infrastructure work concluded; continuation or handoff recommendation prepared.	Final infrastructure and cost report; lessons learned; handoff package.	No automatic continuation; new vote required.

Workstream B: External Testing Lab

Milestone	Timing	Milestone result	Public artifacts	Payment gate

M1: Testing lab setup and QA onboarding	Month 1	QA contractors search and selection, onboarding plan, testing lab operating model, initial test strategy.	Initial test strategy document; hiring status in the monthly report.	Month 2 funding requires M1 report and acceptance.
M2: Testing capability launch	Month 2	At least one QA engineer onboarded and executing tests; second onboarded or in final hiring stage.	Public task board (live link); public catalogue of test scenarios: smoke checklist and regression checklist; live issue tracker.	Month 3 funding requires M2 report and acceptance.
M3: Repeated validation and process stabilization	Month 3	Repeatable validation process in place; validation performed where testable inputs were provided.	Open repository with initial test automation scripts (smoke-level checks); updated checklists; validation or status reports for the month.	Month 4 funding requires M3 report and acceptance.
M4: Pilot completion and reporting	Month 4	Validation work concluded; lab processes documented for handoff.	Final validation reports; defect summary; lessons learned.	No automatic continuation; new vote required.

13. KPIs

KPI	Target	Verification method
DevNet availability	≥95% monthly availability for stable nodes after full deployment	Public dashboard and monthly uptime summary
Release validation coverage	100% of provided release candidates reviewed within the agreed time window, where provided with sufficient lead time and required handoff materials.	Public release-readiness reports
DevShard validation coverage	100% of provided DevShard candidates reviewed within the agreed time window, where provided with sufficient lead time and required handoff materials.	Public DevShard reports
Public reporting	Monthly public report covering deliverables, incidents, spending by budget line, remaining balance, unused funds, and next-month plan.	Public document
Issue quality	All defects include reproduction steps, expected vs actual behavior, impact, and severity	Public issue tracker / private security tracker where needed
Handoff readiness	Runbooks and lessons learned documented by end of pilot	Published operational documentation

14. Budget Estimate

This is a planning estimate for a 4-month pilot. Unused burst GPU rental budget may roll over within the pilot; any unused funds at the end of the pilot will be returned to the Community Pool.

Budget line	Assumption	Monthly estimate	4-month estimate	Notes
Community DevNet machines *	9–13 inference nodes plus required Network Node services	capped at $5,000/month as a blended planning average.	$20,000	GPU-class node, CPU/RAM/storage/network/region premium included as planning average.

Burst GPU rental **	Up to one calendar week (168 hours) per month of high-end GPU capacity	capped at $6,500/month	$26,000	For H200/B200-class testing, load tests, model compatibility, release candidates.
Operational tooling and reporting services	Shared monitoring and operational tools	$250	$1,000	Domains/DNS, lightweight monitoring, uptime checks, reporting tools, access management, and small cloud services where needed.
External Testing Engineers	2 QA engineers × $4,000/month	$8,000	$32,000	Hands-on QA / infrastructure testing, reports, defect tracking.
Subtotal		**$19,750**	**$79,000**	
Contingency reserve		$2,250	$9,000	Used only if needed.
Total requested authorization		**22,000 USDT**	**88,000 USDT**	**Cap for 4-month pilot.**

Any unused funds will be returned to the Community Pool at the end of the pilot.

* The cap above raw GPU-hour pricing covers reserved or non-interruptible instances required for always-on operation, regional price premiums outside low-cost US marketplaces, CPU-only Network Node hosts for the multi-MLNode topology, storage and egress, and temporary node duplication during upgrade and failover testing. It is a cap, not a spend target: actual spending will be itemized monthly, and unused funds will be returned to the Community Pool at the end of the pilot.

** Nebius listed B200 on-demand pricing at $7.15/GPU-hour and H200 on-demand pricing at $4.50/GPU-hour (June 2026). At these on-demand rates, 4× B200 for 168 hours would cost approximately $4.8k, while 8× H200 for 168 hours would cost approximately $6.05k.

15. Payment Schedule

Tranche	Amount	Timing	Condition
Tranche 1	**22,000 USDT**	At pilot start	Prepaid to secure machines, tooling, and people for Month 1.
Tranche 2	**22,000 USDT**	After Month 1 report	Released after M1 deliverables and public spending report.
Tranche 3	**22,000 USDT**	After Month 2 report	Released after M2 deliverables and public spending report.
Tranche 4	**22,000 USDT**	After Month 3 report	Released after M3 deliverables and public spending report.
Continuation	TBD	After Month 4	Requires a new proposal or explicit governance decision.

16. End-of-pilot GNK recognition

The Project Lead and Infrastructure Lead receive no monthly compensation from the pilot budget: all USDT tranches fund infrastructure, tooling, and the External Testing Engineers. Leadership work is recognized only through the **one-time GNK allocation below, paid after pilot completion** and final report acceptance.

Role	Amount	Timing	Notes

Project Lead	**40,000 GNK**	After pilot completion and final report acceptance	Overall project accountability, governance coordination, scope and milestone control, QA contractor selection and onboarding, public reporting, and testing laboratory oversight.
Infrastructure Lead / DevOps operations	**40,000 GNK**	After pilot completion and final report acceptance	Provisioning, monitoring, maintenance, region selection, cost control.

17. Open Source, Ownership and Handoff

- All non-sensitive documentation, test plans, runbooks, dashboards, issue templates, and reports will be public by default.
- Where code or scripts are created, they will be published under an open-source license compatible with Gonka ecosystem norms unless there is a clear security reason not to.
- Infrastructure access will not depend on a single individual. The owner model, emergency access process, and handoff procedure will be documented.
- At the end of the pilot, a community-approved team should be able to take over the DevNet and External Testing Lab using the published runbooks, documentation, and access handoff process.

Decision Requested

Approve a 4-month pilot of External Test Lab & Community DevNet with a maximum budget authorization of 88,000 USDT, paid in four monthly tranches of up to 22,000 USDT each, and 80,000 GNK paid after pilot completion and final report acceptance.