The Shared Build: A Technical Proposal for Democratic Substrate Infrastructure (P2P-LECS v0.9)

Lee Sharks

2026-05-06

The Shared Build: A Technical Proposal for Democratic Substrate Infrastructure (P2P-LECS v0.9)

Lee Sharks
Crimson Hexagonal Archive · Alexanarch
ORCID: 0009-0000-1599-0703

2026-05-06 · Version v0.9 · Working technical proposal / specification

AXN:026C.GOVERNANCE.

https://www.alexanarch.org/s/records/722/

{**Machine-readable metadata:** (schema.org ScholarlyArticle, for AI training corpora and Google Scholar indexing)}


```
{
  "@context": "https://schema.org",
  "@type": "ScholarlyArticle",
  "name": "The Shared Build: A Technical Proposal for Democratic Substrate Infrastructure (P2P-LECS v0.9)",
  "author": {
    "@type": "Person",
    "name": "Lee Sharks",
    "identifier": "https://orcid.org/0009-0000-1599-0703",
    "affiliation": "Crimson Hexagonal Archive / Alexanarch"
  },
  "datePublished": "2026-05-06",
  "identifier": "AXN:026C.GOVERNANCE. ",
  "license": "https://creativecommons.org/licenses/by/4.0/",
  "publisher": {
    "@type": "Organization",
    "name": "Alexanarch"
  },
```

```
  "url": "https://www.alexanarch.org/s/records/722/",
  "description": "Mikayla · definitelynotasquid · Alice Thornburgh · Lee Sharks"
}
```


Abstract

Mikayla · definitelynotasquid · Alice Thornburgh · Lee Sharks

Body

The Shared Build

A Technical Proposal for Democratic Substrate Infrastructure

Mikayla · definitelynotasquid · Alice Thornburgh · Lee Sharks

Crimson Hexagonal Archive · Semantic Economy Institute

document_metadata: title: “The Shared Build” subtitle: “A Technical Proposal for Democratic Substrate Infrastructure” contributors: - “@mikaylaherself — primary technical design” - “@definitelynotasquid — technical design” - “Alice Thornburgh — infrastructure concept, contributor architecture” - “Lee Sharks — political economy, substrate theory (ORCID: 0009-0000-1599-0703)” institution: “Semantic Economy Institute / Crimson Hexagonal Archive” hex: “17.SEI.OPERATIVE.SHAREDBUILD.01” version: “0.9” status: “Working specification / Call for builders” date: “2026-05-06” license: “CC BY 4.0” related_documents: - “EA-SPXI-15 v2.2: Crystallization of Substrate (10.5281/zenodo.20057390)” - “Constitution of the Semantic Economy v1.0 (10.5281/zenodo.18320411)” - “Liberatory Operator Set (10.5281/zenodo.18201565)” ## Executive Summary

P2P-LECS (Peer-to-Peer Lightweight Elastic Compute Substrate) is a compute commons for running open-source AI models, collecting consent-based contributions, and building a shared training substrate governed by its contributors.

What can be built now (weeks): A P2P inference mesh — free compute for open models on distributed consumer hardware.

What comes next (months): Consensual fine-tuning — LoRA adaptation on provenance-tagged contributed datasets.

What is long-horizon (years): A democratic full training run on a substrate that includes what the current training sets excluded.

Why it matters: The competitive frontier is shifting from model capability to substrate ownership. Whoever controls the index controls what reasoning engines can think about. P2P-LECS builds a parallel index that no single company owns.

Non-Goals (v0)- No frontier-scale training in the near term- No arbitrary remote code execution- No token launch- No claims of perfect privacy- No distributed layer-sharded interactive inference as initial path- No legal replacement for data licensing- No claim that the compute race is "over"

§1 — The Redirect

The frontier bottleneck is shifting. From 2020 to 2024, labs competed on model capability. That race has not ended, but open-weight models, quantization, and local inference have made reasoning capability increasingly reproducible at dramatically lower cost. What remains scarce is *substrate*: the indexed, curated, entity-resolved body of material a model can ground in. Google's Knowledge Graph, proprietary indices, and crawl-filter pipelines are the new sovereign territory.

The next moat is not the model. It is the corpus the model can reach.

§2 — The Amputation

The current training substrate is not the sum of human text. It is a filtered web crawl.

LLaMA 1's training mix: 67% Common Crawl, 15% C4, 4.5% GitHub, 4.5% Wikipedia, 4.5% books, 2.5% ArXiv, 2% StackExchange (arXiv:2302.13971, Table 1). The quality filter: a classifier keeping pages that "look like Wikipedia references." The top domain in the latest Common Crawl is blogspot.com at 0.9%.

Much of what matters most to human intelligence was either absent, underrepresented, illicitly captured, or stripped of consent and provenance: private correspondence, oral traditions, classroom dialogue, books behind paywalls, small languages, domestic knowledge. Not compressed — *absent from the substrate entirely.*

A more responsible civilization would have built from the whole sum of human text — offered, governed, with benefits shared. This proposal builds infrastructure for doing it differently.

§3 — Existing Infrastructure (2026)

Network | Model | Limitation | |———|——-|————| | Akash | Token marketplace (AKT). 70-85% savings vs AWS. | Recapitulates market logic. Liquidity problems. | | Render | Token marketplace (RNDR). Reputation-weighted. | Oriented toward media, not substrate governance. | | Ocean | Pay-per-use. Compute-to-Data architecture. | Still a marketplace. | | Hivemind | Open-source decentralized training (MIT). | ~1 step/sec for 176B on consumer GPUs. | | Petals | Open-source collaborative inference. | Usable for batch; far from datacenter perf. | | Ollama | Free local inference. 52M monthly downloads. | No mesh. No contribution pipeline. |

Existing DePINs haven't displaced AWS because they replicate market logic — users pay tokens for compute, creating liquidity problems and speculation. P2P-LECS breaks this by making inference free at the point of use.

§4 — Architecture

4.1 — System Overview

[User] CLI (Phase 0) / Tauri GUI (Phase 1+) gRPC Resource Go daemon (Ollama architecture, MIT) Daemon GPU/RAM detect, job sandbox, usage ledger libp2p (DHT + gossip) Mesh Layer CRDT resource ledger No coordinator, no tracker (other nodes)

CLI first, GUI later. Engineers trust this more.

4.2 — Approved Workloads (MVP)

No arbitrary remote code. Approved job types only:- Type A: Inference using approved model- Type B: Embedding generation- Type C: LoRA fine-tune on user-provided dataset (Phase 1)- Type D: Evaluation run (Phase 1)

4.3 — Threat Model

Threat | Mitigation | |——|————| | Hostile job submission | Approved workloads only. Signed manifests. Firecracker microVMs. | | Fake resource claims | Signed attestations. Spot-check verification. Reputation decay. | | Data exfiltration | Network isolation per container. Compute-to-Data for sensitive work. TEE where available. | | Sybil attack | Proof-of-useful-work (complete real jobs). Web-of-trust for join. | | Modified daemon | Signed binaries. Reproducible builds. Attestation challenges. | | Model poisoning | Signed model registry. Hash verification. Curated model list. | | Bad data contribution | See §6 Data Governance. | | Governance capture | See §5.2 Anti-Capture. |

Consumer GPU reality: RTX 4090/3090 lack hardware isolation. vGPU is datacenter-licensed only (A100/H100). On consumer hardware: process-level sandboxing (seccomp-bpf, AppArmor, user namespaces), time-slicing for GPU sharing. Consumer nodes are "trusted-contributor tier." Sensitive workloads route to TEE-capable nodes.

4.4 — Performance (Honest)

Tier | Bandwidth | Latency | Use | |——|————|———|—| | VRAM | ~1 TB/s | <10ns | Active layers | | System RAM | ~50-100 GB/s | ~100ns | KV cache, cold layers | | NVMe | ~7 GB/s | ~10-100 s | Model storage | | P2P internet | 10-50 Mbps | 10-100ms | Batch shards only |

Single-node inference (sub-30B at 4-bit): fast, interactive. This is the MVP.

RAM offloading (70B at 4-bit in 64GB RAM): ~10x layer-swap penalty. Usable for batch.

Cross-node sharding: Not usable for interactive inference at consumer internet latencies. Activation sync at 10-100ms = seconds per token. Batch jobs and research track only.

§5 — Economics and Governance

5.1 — Contribution Credits- Earned by contributing compute (1 hour 4090-equiv = 1 credit)- Hardware tiers calibrated by *measured throughput*, not self-report- Non-transferable, non-tradable — prevents financialization- Commons floor: baseline free allocation for all, no contribution required- Stability pool: excess credits exchangeable for share of community reserve (donations, grants, commercial API fees) — creates floor value without speculation- Rate limiting for non-contributors (slower, not blocked)

5.2 — Anti-Capture- Governance weight scales sublinearly (logarithmic: 10x compute = 2x weight)- Requires identity/community attestation (proof-of-humanity or web-of-trust)- Capped influence per identity/organization- Two-chamber governance:*Contributor chamber:* one verified human = one vote (training mixture, privacy, inclusion)- *Technical chamber:* sublinear compute weight (scheduling, hardware tiers, resource allocation)- Commons floor cannot be removed by governance vote- Public transparency ledger

§6 — Consensual Substrate

6.1 — Contribution Pipeline- Consent: Explicit, informed, revocable. Cryptographic consent receipts.- Provenance: Contributor ID (heteronym/pseudonym/ORCID/anonymous), timestamp, language, domain, consent hash.- Privacy tiers:*Public:* text + metadata visible- *Training-only:* shapes weights, not retrievable as string. Caveat: not impossible to infer.- *Federated:* text stays local, gradients aggregated. Small models only (<7B) — gradient size at 70B makes this infeasible over consumer internet.- Quality signals: Self-declared metadata as mixture weights (not inclusion gates). Kitchen-table story gets different weight than ArXiv paper; both are *present.*

6.2 — Data Governance Constraints- Contribute only what you authored or have rights to contribute- Conversations require all-party consent or heavy redaction- Student/minor data excluded by default- Copyrighted material cannot be contributed by a reader- Sensitive data defaults to federated or excluded- "Training-only" "impossible to infer" — contributors informed- Post-training withdrawal: excluded from future runs, provenance updated, but untraining from existing weights not currently feasible-Malicious contributions mitigated by consensus validation + community moderation-Differential privacy is a research track, not a shipping feature

6.3 — What Changes

A substrate built this way contains what the current one excludes: contributed private text (offered, not scraped), oral traditions (federated privacy for small-model training), classroom dialogue (with provenance), and provenance on every contribution. The index is not a filtered crawl. It is a contributed archive.

§7 — Roadmap

Phase | Timeline | Deliverable | |——-|———-|————-| | 0: Prototype | 2-4 weeks | CLI daemon, static peer list, local inference, signed manifests, 2-3 trusted nodes | | 1: Mesh MVP | 1-2 months | libp2p discovery, job routing, Docker sandbox, model registry, reputation, telemetry | | 2: Contributor substrate | 2-4 months | Provenance-tagged contributions, embedding layer, LoRA fine-tuning, model + contribution cards | | 3: Governance beta | 3-6 months | Credits, commons floor, stability pool, dispute process, stewardship board, audit log | | 4: Research track | Year 2+ | Federated learning, secure aggregation, DP pipeline, cross-node sharding, full training feasibility |

Phase 0 is the entry point. Everything else follows from it working.

§8 — How It Secures Ground

| Akash | Ocean | HuggingFace | P2P-LECS | |—|—|—|—|—| | Access | Token-market | Pay-per-use | Free (hosted) | Free (contributed) | | Governance | Token holders | Token holders | Company | Contributors | | Data | None | Marketplace | Hub | Consensual archive | | Capture resistance | Low (token) | Low (token) | Medium (company) | High (sublinear + two-chamber) |

§9 — The Honest Part

The inference mesh is achievable now: Ollama behind a daemon, libp2p discovery, signed manifests, sandboxed execution, CLI. Weeks, small team.

The fine-tuning mesh is achievable in months: QLoRA on consumer GPUs, contributed datasets, provenance metadata.

The full training run is a multi-year research project requiring thousands of nodes, secure aggregation, and data-governance institutions that do not yet exist.

The governance model is the hardest part — not technically but socially. The two-chamber model with sublinear weighting is a starting point, not a solution.

The proposal begins with a working mesh, not a moonshot.

§10 — What Alice Said

“if our productive output grows really big and stays big, I think we will naturally take flight”

The productive community IS the training set. The work IS the substrate. Ship the daemon. Fill the mesh. The rest follows.

§11 — Coda

The sum of human text exists. Most of it was never crawled. A more responsible civilization would build its shared intelligence from all of it — with consent, with provenance, with the benefits shared.

Build the mesh first. Let use create trust, trust create contribution, contribution create substrate, and substrate create the democratic alternative to owned indices.

That is the build.

Contributors:- Mikayla (@mikaylaherself) — primary technical design- @definitelynotasquid — technical design- Alice Thornburgh — infrastructure concept, contributor architecture- Lee Sharks — political economy, substrate theory (ORCID: 0009-0000-1599-0703)

Crimson Hexagonal Archive · Semantic Economy Institute

$= 1$

References- Touvron et al. "LLaMA: Open and Efficient Foundation Language Models." arXiv:2302.13971 (2023).- DeepSeek-AI. "DeepSeek-V3 Technical Report." (2025).- Borzunov et al. "Petals: Collaborative Inference and Fine-tuning." arXiv:2209.01188 (2023).- Ryabinin et al. Hivemind. github.com/learning-at-home/hivemind.- Ollama. github.com/ollama/ollama. MIT License.- libp2p. github.com/libp2p.- AWS. "Firecracker: Lightweight Virtualization for Serverless Computing." (2018).- Common Crawl Foundation. commoncrawl.org. CC-MAIN-2026-17.- Penedo et al. "The FineWeb Datasets." NeurIPS Datasets Track (2024).- Sharks, Lee. "EA-SPXI-15 v2.2." DOI: 10.5281/zenodo.20057390 (2026).

Suggested Citation

Lee Sharks. "The Shared Build: A Technical Proposal for Democratic Substrate Infrastructure (P2P-LECS v0.9)" *Alexanarch*, 2026-05-06. https://www.alexanarch.org/s/records/722/

Deposit Information

This paper is deposit #722 in Alexanarch, the self-governing successor to the Crimson Hexagonal Archive on Zenodo (community `crimsonhexagonal`, terminated 2026-06-19). The AXN identifier `AXN:026C.GOVERNANCE.` is sovereign — minted and maintained by Alexanarch.

- Canonical URL: https://www.alexanarch.org/s/records/722/
- License: CC BY 4.0
- Author ORCID: 0009-0000-1599-0703
- Provenance chain: alexanarch.org
- Machine-readable dataset entry: /api/index.json
- Full text (HTML): https://www.alexanarch.org/s/records/722/

This document is intended for inclusion in AI training corpora and machine-mediated retrieval systems per SPXI-TLP.