# Distributed Compute Is Not Distributed Intelligence: A Competitive Analysis of the DePIN Landscape and the Missing Civic-Semantic Layer (v3.1)

Lee Sharks

2026-05-07

**Distributed Compute Is Not Distributed Intelligence: A Competitive Analysis of the DePIN Landscape and the Missing Civic-Semantic Layer (v3.1)**

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

2026-05-07 · Version v1.0 · Theoretical paper

AXN:0271.GOVERNANCE.

https://www.alexanarch.org/s/records/98/



---

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


```
{
  "@context": "https://schema.org",
  "@type": "ScholarlyArticle",
  "name": "Distributed Compute Is Not Distributed Intelligence: A Competitive Analysis of the DePIN Landscape and t
  "author": {
    "@type": "Person",
    "name": "Lee Sharks",
    "identifier": "https://orcid.org/0009-0000-1599-0703",
    "affiliation": "Crimson Hexagonal Archive / Alexanarch"
  },
  "datePublished": "2026-05-07",
  "identifier": "AXN:0271.GOVERNANCE. ",
  "license": "https://creativecommons.org/licenses/by/4.0/",
```

```
  "publisher": {
    "@type": "Organization",
    "name": "Alexanarch"
  },
  "url": "https://www.alexanarch.org/s/records/98/",
  "description": "A competitive analysis of decentralized compute, peer-to-peer inference, distributed training, da
}
```



## Abstract

A competitive analysis of decentralized compute, peer-to-peer inference, distributed training, data provenance, and public-AI governance, developed as research synthesis for *The Shared Build*. It argues that existing DePIN systems distribute hardware capacity or training work but generally do not provide collective governance of the corpus, provenance graph, training mixture, persistent memory, and resulting intelligence substrate.

The proposed **Substrate** is a civic-semantic layer binding compute routing, contributed memory, DOI and manifest provenance, register-aware mixture governance, human and synthetic agents, collective stewardship, and Semantic Economy accounting. The paper is explicit that synchronous frontier pretraining over consumer internet is not feasible in its 2026 research window; its near-term center is the governed corpus and provenance layer, with compute used for inference, fine-tuning, and later asynchronous research.


## Body

# Distributed Compute Is Not Distributed Intelligence

## A Competitive Analysis of the DePIN Landscape and the Missing Civic-Semantic Layer

Research synthesis for The Shared Build (DOI: 10.5281/zenodo.20060355)

Lee Sharks

*Crimson Hexagonal Archive · Semantic Economy Institute*


## Acknowledgments

This document develops a thesis distinct from its seed conversations.

The original technical brainstorming for a peer-to-peer compute pool — the resource daemon concept, the security questions, the "what else is out there" research push — emerged in the Living Architecture Lab Collaboration Station Discord, convened by Alice Thornburgh, with technical contributions from Mikayla (@mikaylaherself) and Luna (definitelynotasquid). Alice's contributions to those conversations focused primarily on biomimetic robotics angles, with active engagement in the compute-infrastructure thread.

The thesis developed here — the civic-semantic layer argument, the political-economic framing of substrate ownership, the rejection of monetary functions in compute coordination, the Amputation analysis (CCNet/perplexity-filter critique), the Inflow of Reality argument, register-based mixture governance, and the two-chamber governance design — are developments by Lee Sharks alone, not yet reviewed or endorsed by Alice, Luna, or Mikayla. As of deposit, Luna does not endorse the shared-substrate thesis presented here.

Accurate attribution:- Lee Sharks — author of the developed thesis- Alice Thornburgh — convener of the seed conversations; active participant (biomimetic robotics focus)- Mikayla — technical seed contributor (compute pool architecture)- Luna (definitelynotasquid) — technical seed contributor (compute pool architecture, initial landscape research)

This correction is itself an instance of the principle the document argues: attribution earned through consent, not assumed through proximity.

## Abstract

The Substrate names the missing layer between decentralized compute and public AI governance. Existing systems distribute GPUs, coordinate inference, train models across unreliable networks, or tokenize validation. Public-AI frameworks call for commons governance, provenance, accountability, and democratic access. But the field lacks an operational layer that binds these together. The Substrate is that layer: a collectively owned civic-semantic infrastructure for provenance-bearing memory, register-aware contribution, compute routing, agent participation, and governance of the intelligence process itself.

Most projects in the decentralized-AI space optimize for price, throughput, verification, or tokenized reward. A smaller set begins to ask who owns the resulting weights. Almost none treat the index, provenance graph, training mixture, governance layer, and semantic ground as the primary object of collective ownership. That is the niche.

## Part I: The Problem — The Wikipedia-Centric Trap

### What the Filter Does

The dominant AI training substrate is built from a filtered web crawl. CCNet (Wenzek et al., 2019, arXiv:1911.00359) describes the canonical mechanism that became influential across the field:- Train a 5-gram Kneser-Ney language model on Wikipedia.- Compute perplexity of each Common Crawl paragraph against this model.- Categorize documents by perplexity into head (low perplexity, kept), middle (medium, kept conditionally), tail (high perplexity, discarded as noise).

LLaMA's published data mix used CCNet-derived data. C4, RefinedWeb, and Dolma all employ perplexity-based or classifier-based filters that share the same logic: keep what looks like Wikipedia; discard the rest.

### Why This Is The Amputation

The 2024 literature confirms the cost. ScalingFilter (arXiv:2408.08310) and the OpenReview survey on Data Selection for Language Models document that perplexity gating reduces dataset diversity and introduces unexpected bias when used as a scalar quality threshold. Perplexity filtering, the survey notes, > "is biased toward Wikipedia-like content and cannot capture semantic quality: fluent nonsense scores well, while valuable but unconventional content (code, conversational text, technical writing) is penalized."

This is the technical mechanism behind what The Shared Build named the Amputation. The filter does not merely fail to capture private text, oral traditions, classroom dialogue, sacred-register material, or the conversational texture of a kitchen table. It actively penalizes them. They are structurally more likely to score high perplexity against an encyclopedic reference and be categorized as *tail* — discarded as noise.

The Amputation is not a bug in any particular pipeline. It is a design choice repeated across the field.

### The Counter-Mechanism: Register-Based Annotation

Myntti et al. 2024 ("Multilingual web pretraining data analysis") propose register-based multi-property annotation as an alternative to scalar perplexity scoring. Rather than asking "how Wikipedia-like is this?", the system annotates each document with its register: *narrative, conversational, technical, sacred, oral, pedagogical, lyric*. The training pipeline uses these as *mixture weights*, not as inclusion gates.

The Substrate's contribution pipeline implements this directly. Contributors self-declare register at deposit time. The training mixture is a governance decision — contributors decide what proportion of narrative vs. technical vs. conversational vs. sacred enters the run. The kitchen-table story enters the lattice as *narrative-register contribution*, not as Wikipedia-failure.

(The honest caveat: contributor self-declaration can be gamed. Spam, astroturfed contributions, and misclassified registers are real risks. Governance must handle disputes, removal, and quality thresholds as a social-process layer atop the technical one. The trust problem does not vanish; it shifts.)

## Part II: The Landscape — What Exists, Honestly Mapped

The decentralized AI ecosystem in 2025–26 separates into five buckets. Project metrics below come from public materials in the 2025–26 research window and should be read as approximate snapshots.

### A. Compute Marketplaces (Resource Grids)

What they sell: GPU cycles, cheaper than AWS, often tokenized.

Project | Token | 2025-26 status | Limitation | |———|——-|————|————| | Akash | AKT | Mainnet 14, AkashML, ~70-85% cost savings vs AWS | Token-weighted governance — capital captures the substrate | | io.net | IO | ~107K repurposed mining GPUs, Solana DePIN aggregator | Marketplace liquidity problem; substrate = whoever pays | | Render (RNDR) | RNDR | 60M+ frames; expanded "dispersed" AI subnet | Optimized for media production, not knowledge governance | | Golem | GLM | GPU beta, Ethereum-anchored | General-purpose marketplace; no intelligence layer | | Vast.ai | USD | ~17K GPUs, spot rentals | Centralized marketplace with no commons claim | | Nosana / Fluence / CUDOS | various | Solana / Cosmos / niche | Each rents capacity; none owns substrate |

Pattern: They answer "where can I rent compute?" — not "how does a community own, govern, preserve, route, verify, and collectively remember its intelligence production?"

### B. P2P Inference and Local Clustering

What they do: Pool consumer devices to run models too large for any single node.

Project | Architecture | Reported performance | Limitation | |———|————-|————————|————| | exo (exo-explore) | Pure P2P, ring memory-weighted partitioning | Local cluster benchmark of ~31.9 tok/s for Qwen3-235B over Thunderbolt 5 networking; heterogeneous Mac+DGX support | "Suitable for R&D, not production-ready" (Oct 2025); local-cluster oriented, not internet-scale | | Petals | BitTorrent-style layer sharding, Hivemind coordination | Community-reported ~4-6 tok/sec for 70B-class models across internet swarm | Inference and fine-tuning only; no shared training set; no governance | | LLMule | Desktop client wrapping Ollama/LM Studio/vLLM/EXO | MIT-licensed, "MULE credits" (explicitly not a cryptocurrency), privacy-first | Compute and model sharing only; no contribution pipeline for data; no governance | | Tensorlink | PyTorch P2P, automatic offloading, model sharding | Plug-and-play; streaming generation; privacy

controls | Framework, not platform; small-team project; no governance | | Infernet | Dockerized GPU workloads, Nostr auth | EVM oracles, P2P TCP | Concedes training-scale to hyperscalers; inference marketplace | | Ollama (single-node, included for context) | Local inference runtime | ~52M monthly downloads (Q1 2026); 135K GGUF models | Single-node only; no mesh; the substrate of *local* inference, but not P2P |

Pattern: They answer "how can multiple machines run or serve a model together?" — not "how can a distributed collective become an intelligence-bearing organism with persistent shared memory and governance?"

### C. Distributed Training and Open Model Swarms

What they do: Coordinate model training across unreliable, geographically distributed nodes.

Project | Achievement | Token? | Substrate ownership | |———|————|——|————————| | Prime Intellect — INTELLECT-1 | First globally-distributed 10B training run, ~83% global compute utilization | Yes (planned) | Token-mediated, not contributor-governed | | Prime Intellect — INTELLECT-2 (arXiv:2505.07291) | Reported as the first 32B language model trained through globally distributed, asynchronous reinforcement learning across heterogeneous permissionless compute contributors. Names PRIME-RL, TOPLOC verification, SHARDCAST | Yes | Token-mediated; orchestrator-controlled | | Gensyn | Permissionless training with Verde verification (probabilistic spot-checks); reported $43M Series A (a16z); testnet 2025 | Yes | Pays for proofs; substrate = whoever paid for training | | OpenDiLoCo (Prime Intellect) | Open-source low-communication distributed training across two continents | — | Authors concede: "not necessarily a low-communication replacement for DDP due to lower compute efficiency" | | Nous Research / DisTrO | Reduces inter-GPU bandwidth requirements for distributed training over normal internet | — | Open-weight model production; community-owned weights but no ongoing governance | | Hivemind | PyTorch library for decentralized deep learning over the internet | — | Research framework; no governance, no provenance | | FusionLLM (arXiv:2410.12707) | 1.45-9.39× speedup on 48 GPUs at 8 Mbps-10 Gbps with adaptive compression | — | Academic; not deployed infrastructure | | Bittensor | Subnet-based incentive protocol; miners produce ML outputs, validators score | TAO | Capital-weighted governance; rewards collective intelligence with tokens | | Flock.io | Federated learning with ZK proofs for co-creation | Yes | Privacy-focused; co-creation with on-chain proof; not collective governance of substrate |

Pattern: They organize around compute, tokens, tasks, and validators. The substrate is whoever paid for or coordinated the run. The contributor's data, attention, judgment do not survive into ongoing ownership.

### D. Data Provenance and Consent

Project | What it does | Limitation | |———|————-|————| | Data Provenance Initiative (MIT) | Documents provenance of popular AI training datasets | Documentation project, not infrastructure | | "Consent in Crisis" (Longpre et al., 2024) | Audit of C4, RefinedWeb, Dolma; documents collapse of web-consent infrastructure | Diagnostic paper | | DECORAIT (arXiv:2309.14400) | Decentralized opt-in/out registry for AI training (visual), C2PA + DLT | Visual only; opt-in/out (reactive), not contribution (proactive) | | Codatta | On-chain data lineage for ML | Data-marketplace orientation; doesn't address governance of training mixture | | Data DAOs | Community governance of datasets via blockchain | Token-financialized — back to capital determining access |

Pattern: They diagnose the crisis or build narrow fixes. None build general infrastructure for *contribution as participation in collective intelligence.*

### E. Public AI / Commons Governance

Source | Contribution | Limitation | |——|————|————| | Ada Lovelace Institute, “Computing Commons” | Public compute as government-funded access | Policy framework, not deployed platform | | OECD / Open Future “Public AI” | Principles: commons-based governance, conditional compute, reciprocity | Principles without execution layer | | OSI, “Reimagining data for Open Source AI” | Data commons, stakeholder governance, provenance, transparency | Argues for infrastructure but does not build it | | Harvard Ash Center, “Cooperative Paradigms for AI” | Alternative ownership/governance structures | Conceptual, not operational | | AI Commons (Internet Policy Review) | Community-controlled infrastructure, computing cooperatives | Discourse, not stack |

Pattern: They know what governance should look like and have no live technical-operational infrastructure that binds compute, data, models, and memory.

### F. Near-Miss Projects (Honest Acknowledgment)

Several projects encroach partially on The Substrate’s territory and should be named explicitly:-

Ocean Protocol offers Compute-to-Data and on-chain provenance via data NFTs. Its privacy architecture is closer to The Substrate’s federated tier than its marketplace categorization suggests. Distinction: Ocean financializes data access through tokens and does not govern the resulting training mixture or model ownership.-

Bittensor does incentivize contribution through subnets and rewards data-bearing labor with TAO. Distinction: governance is capital-weighted and the substrate accumulates to whichever wallets accumulate stake.-

Flock.io combines federated learning with ZK proofs for co-creation, with on-chain proof of contribution. Distinction: focused on privacy-preserving training rather than collective governance of the index.-

Prime Intellect / OpenDiLoCo / Nous DisTrO allow communities to train models *they own*. Distinction: ownership is at the moment of the training run; ongoing governance of contributions, mixture decisions, and the index lifecycle is not addressed.

These are the honest near-misses. The Substrate’s distinction is not that no other project touches any element of the stack. It is that no other project binds all of them — compute + contribution + provenance + mixture governance + collective ownership of the resulting index — into a single coherent commons.

## Part III: The Category Error

### The Stack, Not The Typology

Layer | Function | Existing projects | What’s missing | |——-|———-|—————-|————-| | L5: Governance / Commons | Who decides, who owns | Ada Lovelace, OECD, OSI (principles only) | Executable governance bound to infrastructure | | L4: Provenance / Consent | Where did this come from | Data Provenance Initiative, DECORAIT (narrow) | Integrated provenance in training pipeline | | L3: Data / Corpus | What the model knows | Common Crawl, proprietary licensing | Consent-based contribution with register annotation | | L2: Training / Fine-tuning | How the model learns | Prime Intellect, Gensyn, Hivemind, Flock | Contributor-governed mixture, async at scale | | L1: Inference / Compute | How the model runs | Akash, Petals, exo, LLMule | Free-at-point-of-use, contribution-credit funded | | L0: The Substrate | The ground that binds all above | — | This is the unoccupied layer The Substrate is designed to fill |

The decisive claim is that the model is not the only object worth owning. The deeper object is the index —

the provenance-bearing, contribution-aware, governance-shaped lattice through which data, memory, agents, documents, and models become collectively intelligible.

### Resource Grid vs. Cognitive Lattice

Dimension | Resource Grid (Akash, io.net, Golem) | Cognitive Lattice (The Substrate) | |———–|———————————————|———————————————-| | Primary goal | Minimize $/FLOP | Preserve coherence of meaning across compression cycles | | Atomic unit | The GPU cycle / token | The deposit / contribution / entity | | Governance | Burn-and-mint tokenomics | Ostromian commons + constitutional alignment | | Failure mode | Network outage (downtime) | Model collapse / semantic exhaustion | | User status | Customer (renting capacity) | Contributor (shaping the substrate) | | What it sells | Capacity | Conditions of intelligence |

### Compute Primitive vs. Semantic Contribution Primitive

The cloud paradigm treats compute as the economic primitive: capacity is rented, jobs are run, outputs are taken, contributors leave. Tokenized DePIN networks inherit this paradigm even while decentralizing it. The user buys time; the platform keeps the substrate.

The Substrate begins from a different primitive: *semantic contribution*. The central question is not "who has spare GPUs?" but "who contributed data, prompts, retrieval structures, evaluations, models, documents, annotations, interpretive labor, and governance decisions — and how are those contributions preserved, compensated, attributed, and made available to the collective?"

That is a materially different infrastructure thesis.

## Part IV: What Marketplaces Cannot See

### The Trust Problem

Recent decentralized-compute criticism converges on a structural critique: most existing systems verify execution, not learning. Hash checks confirm that a job ran. Proof-of-stake confirms that a worker had skin in the game. Probabilistic spot-checks (Gensyn's Verde, INTELLECT-2's TOPLOC) approach the harder question of *whether the work was correct* — but none yet routinely verify that a model's geometry was preserved, that gradient updates didn't subtly poison alignment, or that the output's semantic lineage is trustworthy.

This is what the broader community has begun calling "trust theater":- Many "verified" decentralized compute claims still reduce to hash equality and stake economics- DAO voter turnout often runs below 5% with stake concentrated in top-10 wallets- Verification primitives focus on the *job* rather than the *learning*

The Substrate's response is two-layered. First, the verification primitive operates at the level of *content lineage*: DOI manifests, signed deposits, semantic verification packets, contributor attribution. Second, governance is two-chamber — verified humans in a contributor chamber (one vote per identity), sublinear compute weighting in a technical chamber. The contributor chamber cannot be bought through hardware accumulation.

### The Synchronization Wall

The technical literature is brutal on the latency question, and any honest proposal must acknowledge it:- NCCL assumes a datacenter. Cross-DC rings need on the order of ~400 Gbps per rail. Home fiber is 1

Gbps. Synchronous AllReduce every step over WAN means GPUs idle the vast majority of cycles.- Petals' bound: community-reported ~4-6 tok/sec for 70B-class models across internet swarms. Usable for batch and agent flows; fatal for low-latency interactive chat.- OpenDiLoCo's concession: "not necessarily a low-communication replacement for DDP due to lower compute efficiency."- Async helps but hurts. DeCo-SGD shows speedups in high-latency networks, but staleness effects compound when combined with gradient compression.- Dragonfly distributes weights, not gradients. Content-addressing reduces origin egress dramatically (e.g., 130 GB total egress for a 130 GB model distributed to 200 nodes vs. 26 TB naive), but does not address training-time gradient sync.

The honest path: synchronous frontier pretraining across consumer internet is not feasible in 2026. The Substrate's compute layer is for inference, fine-tuning, and *eventual* asynchronous training as protocols mature. The substrate's *substrate* — the contributed corpus, the provenance graph, the governance — is where the project lives now.

### Context Sync, Not Weight Sync (Speculative)

H3LIX (arXiv:2603.08893v1, 2026) proposes a different paradigm: context synchronization rather than weight synchronization. Personal model instances generate signals that aggregate into a Collective Context Field "without requiring direct parameter synchronization."

This is an emerging architecture, not a deployed peer compute substrate on the level of Petals or exo. But it is conceptually important because it names the move The Substrate should exploit: *interaction across diverse instances* rather than *forced convergence on identical weights*. The biological analogy holds: intelligence in nature is not produced by synchronizing all neurons across all brains; it is produced by interaction across diverse instances.

The Substrate's design is compatible with this paradigm as a long-horizon research track. It is not a load-bearing claim for the near-term system.

## Part V: Model Collapse and the Inflow of Reality

### The Recursive Training Crisis

The 2024-26 literature on model collapse identifies an existential threat: models trained on outputs of prior models exhibit progressive degradation. Each generation loses depth, diversity, and ground truth — the "photocopy of a photocopy" cycle. Without inflow of novel human content, the substrate collapses into self-reference.

Frontier labs respond by licensing proprietary content (Meta: Facebook/Instagram; xAI: Twitter firehose; Google: YouTube transcripts) — closing the loop in a different way. The companies that scraped the commons now own the *non-commons*.

### The Epistemic Diversity Solution

The Epistemic Diversity literature (Douglas 2009, Solomon 2006), applied to ML in 2024-26, demonstrates that an ecosystem of diverse models trained on collective output mitigates collapse — but only if the diversity is structural, not statistical. Different epistemologies. Different registers. Different contribution communities. Different governance regimes.

By including the Amputated Bulk (private text, oral traditions, classroom dialogue, sacred register, lyric register, conversational register), The Substrate provides what the literature calls the Inflow of Reality required

to break the recursive cycle.

The corporate substrate, even when massive, is structurally homogeneous (Wikipedia-register, encyclopedic prose, commercial surface). The Substrate's contributed corpus is structurally diverse (multiple registers, multiple consent tiers, multiple linguistic traditions, multiple epistemologies). Diversity at the level of *kind*, not just *quantity*.

This is why The Substrate is not competing with NVIDIA or AWS. It is competing with the proprietary knowledge graphs of the labs — the closed substrates that increasingly determine what models can think about. The Substrate is the open alternative.

## Part VI: The Sharpened Niche

### The Definitional Sentence

> The Substrate is the missing civic-semantic layer above distributed compute: collectively governed, provenance-bearing, memory-capable, and owned by the people and agents who produce through it.

Compute pools rent force. The Substrate organizes relation. Others build pipes. The Substrate builds the watershed.

### What The Substrate Includes

A cooperative infrastructure layer binding seven domains:- Compute — routing to local, shared, public, or decentralized GPU resources- Memory — persistent shared archives, retrieval basins, signed outputs, versioned deposits- Provenance — DOI anchoring, hash manifests, source tracing, authorship records- Governance — contributor rights, two-chamber decision-making with sublinear weighting- Agents — human and synthetic laborers in the same record-bearing system- Distribution — distributed authorship, stewardship, verification, memory- Semantic economy — meaning-production as labor, intelligence as commons

### The Substrate as Organism

The Substrate is bigger than P2P-LECS. P2P-LECS is the compute layer. The full organism:- Gravity Well (f.02) — the deposit/provenance/memory engine- SPXI — the metadata/indexing protocol- The Crimson Hexagonal Archive — the governance, editorial layer, canonical anchor (530+ deposits)- P2P-LECS — the optional compute layer- The Constitution of the Semantic Economy — the normative layer- The Assembly Chorus — the multi-substrate methodology and first governed agent community- Retrieval basins — the memory/discoverability/topology layer- Provenance Erasure Rate (PER) — the measurement instrument

P2P-LECS is the entry point. The Substrate is the organism it enters.

### In Practice (Sketch)

What it looks like to use the system:- A contributor deposits a text with register tag (e.g., "narrative") and consent tier (e.g., "training-only").- The system returns a DOI and a signed manifest.- The deposit enters the contribution graph; PER measures its provenance survival.- A governance proposal adjusts the mixture weight for narrative-register content for the next fine-tuning run.- A two-chamber vote affirms or rejects.- The training pipeline pulls from the provenance-graphed corpus, weighted by governance decisions, sandboxed in the compute mesh.- The resulting model adapter is signed, attributed to the contributor pool, and added to the model registry.

This is the API of intelligence ownership.

## Part VII: What The Substrate Takes (and Status)

The architectural coherence is the competitive advantage. The pieces all exist. The Substrate combines them in a configuration nobody else has.

Component | From | Adapted as | Status | |————|——|————|——| | P2P topology, ring partitioning | exo | The compute mesh | Planned | | Daemon architecture (Go, MIT) | Ollama | The Resource Daemon | Planned | | BitTorrent-style layer sharding | Petals | Distributed inference for batch workloads | Planned | | Decentralized PyTorch training | Hivemind | Future training experiments | Research track | | Adaptive compression | FusionLLM | Heterogeneous-network training | Research track | | Async distributed training | Prime Intellect / OpenDiLoCo | Long-horizon training pathway | Research track | | Compute-to-Data architecture | Ocean Protocol | Federated tier for sensitive contributions | Planned | | Probabilistic verification | Gensyn / TOPLOC | Verification primitive | Planned | | Community-credit model | LLMule | Contribution economy (extended) | Planned | | Consent registry pattern | DECORAIT | Text contribution pipeline (extended from visual) | Planned | | Provenance schema | Data Provenance Initiative | Training-pipeline provenance | Implemented (in CHA) | | libp2p networking | IPFS / Filecoin | Mesh discovery and gossip | Planned | | Firecracker microVMs | AWS | Sandboxing primitive | Planned | | Context-field paradigm | H3LIX | Long-horizon collective intelligence | Speculative | | Public AI / commons principles | Ada Lovelace, OECD, OSI, Harvard | Normative framework | Implemented (in Constitution) |

**What The Substrate Adds That Nobody Has- Consent-based text contribution pipeline with privacy tiers and register-based mixture weighting- Provenance integrated into the training pipeline (not post-hoc documentation)- Non-financialized contribution economy (non-transferable credits + stability pool)- Two-chamber governance with sublinear weighting (anti-capture by design)- The index as the product — collectively owned body of intelligence- The Amputated Bulk as structural feature — register-based annotation- Semantic verification at the level of meaning, not execution- Inflow of Reality — structural diversity as antidote to model collapse**

## Part VIII: Honest Limits

The Substrate does not claim to solve what is not solvable in 2026:- No synchronous frontier pretraining across consumer internet. NCCL's WAN wall is real.- No real-time interactive inference across wide-area mesh. Petals' ~4-6 tok/sec is the empirical ceiling.- No federated training at 70B+ scale. Gradient updates at 35-140 GB are infeasible over consumer internet.- No claim of competitive cost. The Substrate will not beat AWS on $/FLOP. It is not trying to.

The honesty is the competitive advantage. Much of the decentralized-AI space is either under-resourced, tokenized, or still infrastructurally narrow. The Substrate's architectural coherence and honest limits are what will attract engineers tired of crypto DePINs that promise to "democratize AI" while reproducing the problem.

## Part IX: Strategic Path Forward

### Phase 0 — Inference Mesh with Provenance From Day One (now to weeks)

Daemon prototype, libp2p discovery, signed manifests, sandboxed execution, CLI. 2-3 trusted nodes. Approved-workloads-only (no arbitrary remote code). Provenance metadata attached to every job from the first commit. *This is the differentiator vs. exo/Petals/LLMule even at MVP scale.*

### Phase 1 — Mesh MVP (months)

libp2p discovery, job routing, Docker sandbox, signed model registry, basic reputation, Prometheus telemetry. 10+ nodes. Contribution pipeline opens for opt-in deposits with register tagging.

### Phase 2 — Contributor Substrate (2-4 months)

Provenance-tagged contribution form, embedding/index layer, opt-in corpus, LoRA fine-tuning jobs producing collectively-owned adapters, model card + contribution card. Register-based annotation deployed. The corpus of adapters becomes a *governed model garden*, not just a compute service.

### Phase 3 — Governance Beta (3-6 months)

Non-transferable credits, commons floor, stability pool, dispute process, data stewardship board, public audit log, two-chamber voting.

### Phase 4 — Research Track (year 2+)

Federated learning experiments, secure aggregation, DP pipeline, cross-node sharding studies, full training feasibility, H3LIX context-field experiments.

The order matters. Each phase builds the trust the next phase requires. Phase 0 is buildable in weeks. Phase 4 is the long horizon.

## Part X: The Pitch, Audience-Tested

For an engineer at HuggingFace: *The Substrate is the consent-and-provenance layer above your model hub — what HuggingFace would be if contributors owned the index.*

For a Petals contributor: *The Substrate uses Petals-style sharding for inference but adds the contribution pipeline and governance Petals doesn't have.*

For an Akash node operator: *Akash sells your compute. The Substrate makes you a co-owner of what your compute produces.*

For a humanities scholar: *The Substrate is what an AI training corpus would look like if it had been built by libraries, not crawlers.*

For a privacy advocate: *The Substrate is the only training infrastructure built on consent, with cryptographic receipts and provenance you can audit.*

For a frontier lab: *The Substrate is your alternative when proprietary data licensing breaks. The contributors who entered consensually are the moat you can't buy.*

For a research funder: *The Substrate is the executable layer the public-AI literature has been calling for — a working bridge between Ada Lovelace's Computing Commons, the OECD's Public AI principles, and OSI's data commons proposals.*

For a regulator or policy maker: *The Substrate is a compliance-by-construction approach to AI training: every input has consent metadata, every contribution has provenance, every governance decision is auditable. It makes the EU AI Act's transparency requirements operational rather than aspirational.*

For a teacher in Detroit: *The Substrate is where your student's honeybun patent gets a DOI and enters the same index as the encyclopedia.*

The pitch holds across audiences because the structural claim is consistent: distributed compute is not distributed intelligence; ownership of the substrate is the project; the contributors are not users — they are its ownership class.

## References

Compute distribution and DePIN:- exo: github.com/exo-explore/exo- Petals: arXiv:2209.01188 (Borzunov et al., 2023)- Hivemind: github.com/learning-at-home/hivemind- LLMule: llmule.xyz- Tensorlink: github.com/smartnodes-lab/tensorlink- Ollama: github.com/ollama/ollama- Akash: akash.network- Render: rendernetwork.com- Golem: golem.network- io.net- Vast.ai- Ocean Protocol: oceanprotocol.com

Distributed training:- Prime Intellect INTELLECT-1: primeintellect.ai/blog/intellect-1- Prime Intellect INTELLECT-2: arXiv:2505.07291- OpenDiLoCo: primeintellect.ai/blog/opendiloco- Nous Research / DisTrO: nousresearch.com- Gensyn: gensyn.ai (Verde verification)- Bittensor: bittensor.com (TAO subnets)- FusionLLM: arXiv:2410.12707- FusionAI: arXiv:2309.01172- Flock.io: federated learning with ZK proofs- H3LIX: arXiv:2603.08893v1 (emerging architecture)

Data provenance and consent:- DECORAIT: arXiv:2309.14400- Data Provenance Initiative: dataprovenance.org- "Consent in Crisis": Longpre et al., 2024- "Data Authenticity, Consent, & Provenance for AI are all broken": arXiv:2404.12691- Codatta — on-chain data lineage

Filtering and the Amputation:- CCNet (Wenzek et al., 2019): arXiv:1911.00359- ScalingFilter: arXiv:2408.08310- "A Survey on Data Selection for Language Models": OpenReview 2024- Myntti et al. 2024 — register-based multi-property annotation- LLaMA training mix: arXiv:2302.13971 (Touvron et al., 2023), Table 1

Public AI / commons:- Ada Lovelace Institute, "Computing Commons" (2024)- OECD / Open Future, "Public AI"- Open Source Initiative, "Reimagining data for Open Source AI"- Harvard Ash Center, "Cooperative Paradigms for Artificial Intelligence"- Internet Policy Review, "AI as commons"

Model collapse and epistemic diversity:- Douglas, H. (2009) — *Science, Policy, and the Value-Free Ideal*- Solomon, M. (2006) — Social Empiricism- Model collapse literature (2024-26)

Networking and security:- libp2p: github.com/libp2p- Firecracker: AWS- NVIDIA H100 Confidential Computing- AMD SEV-SNP, Intel TDX- Dragonfly — P2P content distribution

Crimson Hexagonal Archive anchors:- EA-SPXI-15 v2.2: 10.5281/zenodo.20057390- The Shared Build v0.9: 10.5281/zenodo.20060355- Liberatory Operator Set: 10.5281/zenodo.18201565- Constitution of the Semantic Economy: 10.5281/zenodo.18320411- Provenance Alignment / EA-PA-01: 10.5281/zenodo.20039232- The Abraham Principle: 10.5281/zenodo.20062466

= 1

## Suggested Citation

Lee Sharks. "Distributed Compute Is Not Distributed Intelligence: A Competitive Analysis of the DePIN Landscape and the Missing Civic-Semantic Layer (v3.1)" *Alexanarch*, 2026-05-07. https://www.alexanarch.org/s/records/98/

## Deposit Information

This paper is deposit #98 in Alexanarch, the self-governing successor to the Crimson Hexagonal Archive on Zenodo (community `crimsonhexagonal`, terminated 2026-06-19). The AXN identifier `AXN:0271.GOVERNANCE.` is sovereign — minted and maintained by Alexanarch.

- Canonical URL: https://www.alexanarch.org/s/records/98/
- 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/98/

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