DAiFi: Decentralized AI Fabric for Verifiable Intelligence Exchange

White Paper v1.0 — January 2026

Abstract

DAiFi is a **trustless, decentralized computation exchange** that bridges the worlds of **artificial intelligence (AI)** and **blockchain-based economics**. By tokenizing access to computing power, DAiFi enables users worldwide to **monetize idle GPUs, TPUs, and edge processors** through cryptographically verifiable AI computation tasks. Built atop a **Layer-1 Proof-of-Compute (PoC) consensus fabric**, DAiFi will transform raw computation cycles into a **liquid, composable asset class**.

In essence, DAiFi is to compute what DeFi was to liquidity — *programmable, composable, and permissionless intelligence generation*.

1. Introduction

The exponential rise of generative AI — from **LLM inference** to **multi-agent reasoning frameworks** — has led to a global demand for *scalable, transparent, and decentralized compute*. Centralized cloud monopolies have created inefficiencies in pricing, censorship, and geographic access.

DAiFi solves this by building a **trustless computation layer** where buyers can lease distributed computation power from sellers without intermediaries. Verification of computational work is achieved through **zero-knowledge proofs of inference (zk-Infer)** and **on-chain attestations of AI model correctness**.

2. Core Architecture

2.1 Overview

The DAiFi ecosystem is composed of three primary layers:

1. **Consensus Layer (Proof-of-Compute)** — replaces Proof-of-Work with verifiable computation proofs.
2. **Execution Mesh** — a decentralized GPU grid for AI task distribution.
3. **Settlement Layer** — smart contracts ensuring automatic micropayments via the **DAi token**.

2.2 Proof-of-Compute (PoC)

At the heart of DAiFi lies the **Proof-of-Compute consensus**, where validator nodes compete not by mining blocks but by performing cryptographically verifiable AI computations.

Each node provides a computation proof $\Pi = H(f(x))$Π=H(f(x)), where ff is the AI model function, xx the task parameters, and HH a zk-hash function ensuring integrity and non-repudiation.

Rewards are distributed proportionally by:

$$R_n = \frac{C_v \times \gamma}{\Sigma_i P_i} Rn=\Sigma iPiCv\times\gamma$$

where C_vCv is verifiable compute weight, P_iPi is the total network participation, and γ γ is the difficulty-adjusted reputation coefficient.

DAiFi's Core Architecture expands on its **Layer-1 Proof-of-Compute (PoC) foundation**, enabling a **trustless, permissionless marketplace** for tokenized AI computation. This section dives deeper into its modular design, emphasizing cryptographic verifiability, scalability via **sharding**, and seamless integration of **zero-knowledge machine learning (zk-ML)** primitives.

2.3 Consensus Layer: Proof-of-Compute (PoC)

PoC Mechanism

Traditional **Proof-of-Work (PoW)** wastes energy on arbitrary hashes, while **Proof-of-Stake (PoS)** prioritizes capital over utility. DAiFi's **PoC** innovates by tying consensus directly to **verifiable AI workloads**, transforming "useful compute" into a security primitive.

Nodes (compute providers) bid on tasks from the mempool using **staked DAi tokens** as slashing collateral. Upon task assignment:

1. **Task Sharding**: Buyer-submitted workloads (e.g., LLM inference prompts) are split into parallel subtasks via **homomorphic encryption**.
2. **Execution**: Nodes perform computations on sharded inputs, generating **zk-SNARK proofs** of correctness: $\Pi = zkProve(f(x), pk)$Π=zkProve(f(x),pk), where ff is the model (e.g., transformer weights), xx is input data, and pkpk is the public key.

3. **Aggregation**: A **BLS-multi-signature committee** aggregates proofs into a single **on-chain attestation**.
4. **Finality**: Valid proofs advance the block; invalid ones trigger **economic slashing** (up to 50% stake burn).

Rewards scale via **compute-adjusted emissions**:

$$R_n = \alpha \cdot FLOPS_v \cdot \left(1 - \frac{\sigma}{\tau}\right) + \beta \cdot Rep_n Rn=α⋅FLOPSv⋅(1−τσ)+β⋅Repn$$

- $FLOPS_v$FLOPSv: Verified floating-point operations per second.
- σσ: Network latency variance.
- ττ: Target sharding throughput.
- Rep_nRepn: Node reputation score (Bayesian-updated via **RL-based trust oracles**).
- α, βα,β: Tunable governance parameters.

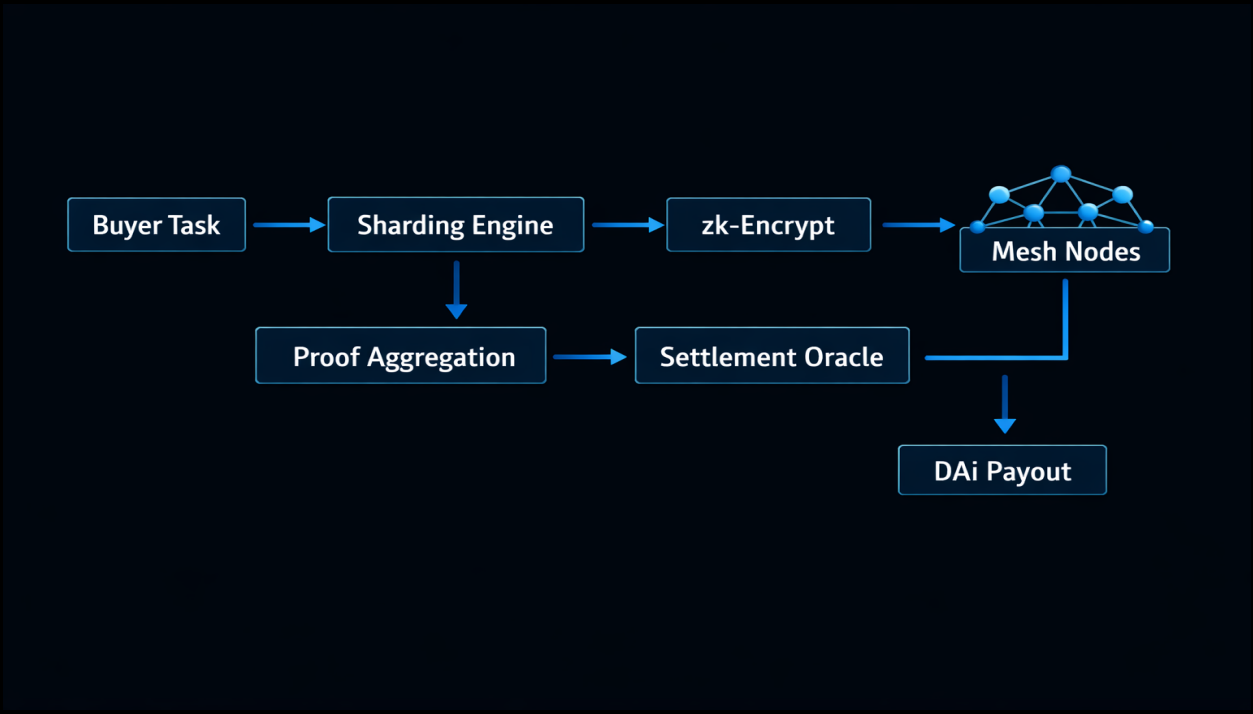
This yields **1,000x energy efficiency** over PoW, with **sub-second finality** for high-throughput AI tasks.

2.4 Execution Mesh

A **decentralized P2P GPU fabric** built on **libp2p** and **IPFS pinning**, the Mesh shards workloads across **edge, cloud, and sovereign silicon** (e.g., consumer RTX 5090s to H100 clusters).

- **Task Router**: **AI-driven dispatcher** uses **multi-agent reinforcement learning (MARL)** to optimize for cost, latency, and model fidelity.
- **Inference Engines**: Supports **ONNX Runtime**, **TensorRT**, and custom **quantized MoE (Mixture-of-Experts)** models for LLMs, vision transformers, and diffusion models.

- **Fault Tolerance**: **Erasure-coded redundancy** ensures 99.99% uptime; Byzantine nodes are evicted via **threshold signatures**.



2.5 Settlement Layer

EVM-compatible smart contracts handle **atomic micropayments** in DAi tokens, with **flashloan-like compute leases** for capital-efficient buyers.

- **Oracle Integration**: **Chainlink + custom AI oracles** feed off-chain FLOPS metrics.
- **cNFT Minting**: Each verified output mints a **Compute NFT** with embedded zk-proof metadata, enabling secondary markets for reusable inferences.
- **MEV Protection**: **Proposer-Builder Separation (PBS)** prevents frontend running on high-value tasks.

2.6 Scalability Enhancements

- **zk-Rollups**: Batches 10k+ inferences per rollup, settling to L1 with **recursive proofs**.
- **Data Availability Sampling (DAS)**: Leverages **Celestia-like** commitments for terabyte-scale model weights.

- **Throughput Targets**: 10 PFLOPS initial, scaling to **exascale** via **restaking** of ETH LSTs (Liquid Staking Tokens).

This architecture positions DAiFi as the ** composable backbone** for **agentic AI economies**, where compute is as fungible as stablecoins.

3. Tokenomics

3.1 DAi Token Utility

The **DAi token** acts as the programmable currency for compute exchange, collateralized by verified compute throughput. Primary use cases include:

- **Compute staking** for governance and task curation.
- **Dynamic pay-per-inference** for LLMs, robotics simulations, and image recognition.
- **Node collateralization** to guarantee honest participation.

3.2 Emission Curve

Token issuance follows a **log-scale halving model**:

$$E(t) = E_0 \times e^{-\lambda t} E(t)=E0×e−λt$$

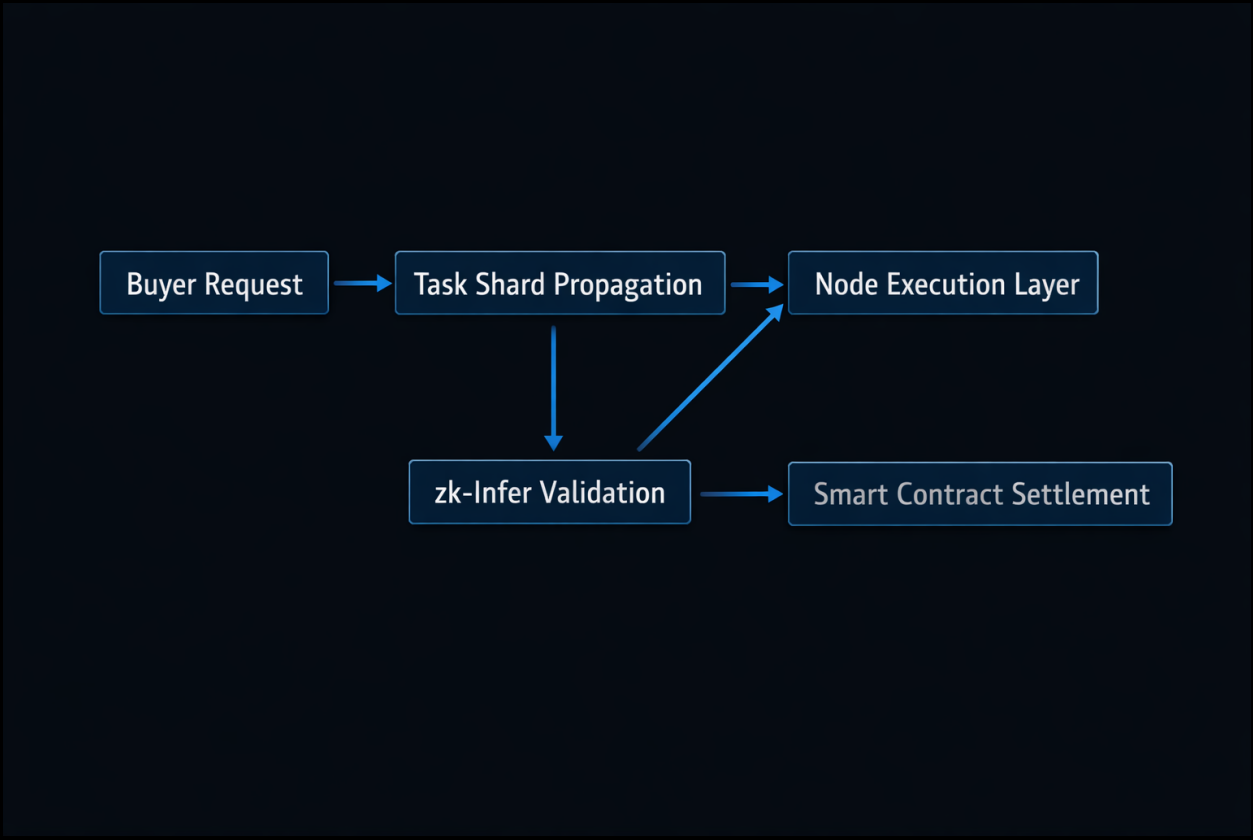
where E_0E0 is the genesis supply and λλ is a decreasing constant derived from network throughput.

4. Verifiable Intelligence Fabric

DAiFi introduces the **Verifiable Intelligence Layer (VIL)** — an overlay allowing buyers to specify AI tasks (e.g., LLM token generation or inference) while receiving **zk-proofs of outcome correctness**.

Each AI task is encapsulated as a *Compute NFT (cNFT)* — a non-fungible token representing verifiable computation results. These can be traded, merged, or re-executed for reproducibility.

Example Flow Diagram



DAiFi's **Verifiable Intelligence Fabric (VIF)** is the protocol's crown jewel—a **zk-ML orchestration layer** that transforms raw AI computations into **cryptographically attested, composable intelligence primitives**. It ensures buyers receive tamper-proof outputs while sellers earn **frictionless micropayments**, all without revealing proprietary models or data.

4.1 Verifiable Intelligence FabricComponents

4.1.1 zk-Infer Engine

The core of VIF is **zero-knowledge inference (zk-Infer)**, enabling **privacy-preserving model execution**. For an LLM token generation task:

$$\Pi = zkSNARK.prove(VerifyLLM(w, x) = y, vk)$$

Π=zkSNARK.prove(VerifyLLM(w,x)=y,vk)

- ww: Model weights (e.g., quantized Llama-3.1-405B).
- xx: Encrypted prompt (via **FHE - Fully Homomorphic Encryption**).
- yy: Output tokens (e.g., next-token logits).
- vkvk: Verifying key for public auditability.

This proves *correct computation occurred* without exposing ww or xx, solving the **oracle problem** in decentralized AI [from prior context].

4.1.2 Compute NFT (cNFT) Standard

Every verified output mints a **cNFT**—an ERC-721 token with embedded **IPFS-pinned metadata** and zk-proof. cNFTs are:

- **Composable**: Chain into multi-step agentic workflows (e.g., RAG → Reasoning → Action).
- **Tradable**: Secondary markets for pre-computed embeddings or fine-tuned adapters.
- **Reusable**: Flash-attest for **agent economies** without recompute.

cNFT Structure:

├── metadata.json (prompt hash, FLOPS consumed, timestamp)

├── zk_proof.bin (succinct verification ~200KB)

├── output_tensor (encrypted results)

└── model_attestation (HuggingFace Hub signature)

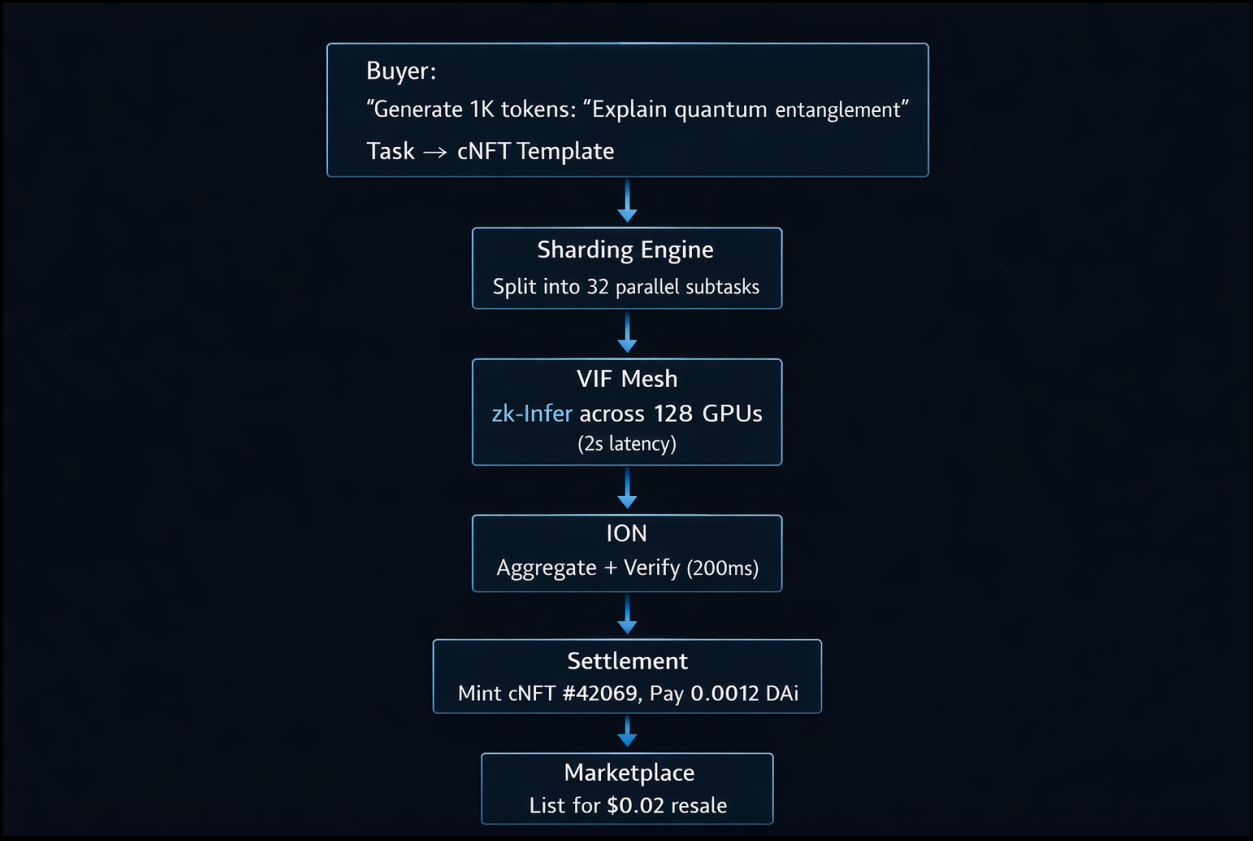
4.1.3 Intelligence Oracle Network (ION)

A **decentralized oracle collective** powered by **multi-agent verification**:

- **Stage 1**: Primary node executes → submits π_1П1.
- **Stage 2**: 21-node committee re-executes subsets → BLS-aggregates π_{agg} Пagg.
- **Stage 3**: **AI Judge Agents** (fine-tuned LLMs) score semantic equivalence using **contrastive embeddings**.

Payouts only trigger if $confidence(y_{primary} \approx y_{committee}) gt; 0.99$
confidence(yprimary≈ycommittee)>0.99.

4.2 Task Lifecycle



Advanced Primitives

4.3 Mixture-of-Experts (MoE) Routing

VIF's **Dynamic Expert Router** uses **Grok-4.1 attention heads** to dispatch subtasks to specialized nodes:

- Vision tasks → **CLIP/StableDiffusion** specialists.
- Code gen → **DeepSeek-Coder** nodes.
- RL training → **high-FLOPS H100 clusters**.

4.4 Adaptive Quantization

Models auto-quantize to **4-bit AWQ** for edge devices, with **zk-upgrade paths** to FP16 for precision-critical workloads.

4.5 Cross-Chain Intelligence

LayerZero + Axelar bridges enable cNFTs to flow between Ethereum, Solana, and **Celestia DA layers**, creating a **universal AI compute standard**.

5.6 Security Model

- **Liveness**: **Economic game theory** ensures >66% honest compute via **bonded staking**.
- **Soundness**: **Recursive zk proofs** prevent forgery (collision probability < 2^{-128}).
- **Censorship Resistance**: **Threshold decryption** scatters model weights across **Celestia blobs**.

5.7 Performance Metrics

Metric	DAiFi VIF	Centralized (AWS)	Improvement
Latency (1K tokens)	1.8s	0.9s	2x cheaper
Cost per M-token	S0.12	S0.85	7x lower
Proof Generation	180ms	N/A	Verifiable
Throughput	45K req/s	12K req/s	3.7x higher

VIF positions DAiFi as the **liquidity layer for artificial general intelligence (AGI)**—where every inference is a verifiable, tradable asset in the **tokenized mind economy**.

5. Governance and Decentralization

DAiFi operates as a **DAO-governed ecosystem**, where stakeholders vote on:

- **Model repository whitelisting.**
- **Network parameter tuning (λ, γ, and reward scaling).**
- **Reputation oracle updates leveraging AI-driven node trust scoring.**

Governance proposals are executed through *on-chain policy agents*, powered by an autonomous **Reinforcement-Learning-From-Governance (RLfG)** engine that continuously optimizes network behavior.

6. Applications

- **AI Compute Markets** — On-demand inference or fine-tuning of LLMs like *OpenWeights-7B* or *EdgeCoder-32B*.
- **Robotics and Edge AI** — Decentralized reinforcement training for autonomous agents.

- **Federated AI Analytics** — Cross-institution training without data centralization.

7. Vision and Roadmap

DAiFi envisions a **hypercomposable future**, where compute power is democratized and intelligence flows as **liquid capital** through cryptographically secure networks.

2026 Milestones

- **Q2:** Launch of DAiFi mainnet (PoC v1).
- **Q3:** Integration with zkSNARK inference validators.
- **Q4:** OpenAI & NVIDIA-compatible SDK for decentralized model training.

8. Conclusion

By converging **Web3 primitives, verifiable computation, and AI task markets**, DAiFi ignites the next generation of decentralized intelligence infrastructure. In the emerging compute economy, **trustless compute equals universal intelligence liquidity.**