



DGrid AI: The Decentralized AI Inference Network for Open, Low-Cost & Community-Powered AI

Litepaper - June 2025

<https://dgrid.ai>

1 Abstract

DGrid is a decentralized AI inference network built on a core philosophy: ***"AI should serve the open network, not centralized platforms."*** We aim to break the monopoly of centralized AI providers, enabling community-driven participation in AI infrastructure and integrating large language model (LLM) capabilities into the open network ecosystem—delivering fair, transparent, and accessible inference services for all users.

DGrid.AI is redefining the future of AI by constructing a decentralized AI inference routing network based on the Proof of Quality (PoQ) mechanism, which enables freer and more efficient flow of AI inference. Unlike traditional centralized AI systems, the DGrid network empowers every node to participate, every call to be traceable, and makes AI a fundamental capability of the blockchain world.

DGrid combines AI RPC, LLM inference, and a distributed node network to solve critical pain points of centralized AI—high costs, uncontrollable services, single points of failure—and fill the gaps in Web3 AI, such as the lack of unified interfaces and trusted inference environments.

Core Values

- **Decentralization:** AI belongs to the open network, not centralized platforms.
- **Accessibility:** Anyone can join as a node operator or developer.
- **Trustless:** Powered by on-chain permissions and auditing mechanisms.
- **Scalability:** Modular design enables flexible deployment of models, inference, and caching.
- **Value Discovery:** Enabling inference resources to be "routeable, verifiable, and comparable in price" through intelligent routing, achieving efficient allocation of intelligent liquidity.

2 Background

In the current AI landscape, both centralized systems and Web3 ecosystems face critical limitations that hinder the seamless integration of AI and blockchain.

2.1 Pain Points of Centralized AI

Centralized AI platforms, despite their dominance, suffer from inherent flaws:

- **Excessive Costs and Opacity:** High expenses for data centers, GPU

maintenance, and model iteration are passed to users, and users struggle to compare the true relationship between inference quality and cost.

- **Unreliable Services and High Barriers:** Dependence on single providers leads to unannounced outages, arbitrary API changes; meanwhile, different models have varying API formats and key systems, resulting in high integration costs for developers—posing risks for Web3 applications requiring 24/7 stability.
- **Censorship and Single Points of Failure:** Centralized control enables content moderation bias, while a single decision or ban can determine the fate of countless AI applications, conflicting with the openness of the internet

2.2 Gaps in Web3 AI Infrastructure

Web3 applications demand AI capabilities (e.g., smart contract analysis, DeFi strategy generation) but lack foundational tools:

- **Fragmented Interfaces:** No universal protocol for LLM interactions, forcing developers to build custom integrations for each model or provider.
- **Untrusted Inference Environments:** Inference results from centralized providers cannot be verified on-chain, risking manipulation in high-stakes scenarios (e.g., oracle feeds).
- **Missing Decentralized Execution Infrastructure:** Without a scalable, community-operated network for running LLM inference, Web3 AI remains conceptual—unable to achieve large-scale adoption.

These challenges create a critical bottleneck: AI cannot effectively serve the Web3 ecosystem without a trustless, standardized, and decentralized inference layer. DGrid.AI was built to address these pain points and fill these gaps.

DGrid.AI's mission is to "reconstruct the underlying structure of AI inference" by building a decentralized infrastructure that frees AI operations and applications from dependence on a few large platforms. Our ultimate goal is to make AI a native, foundational capability of the blockchain world—seamlessly integrated into blockchain applications, just like data storage and transaction processing today.

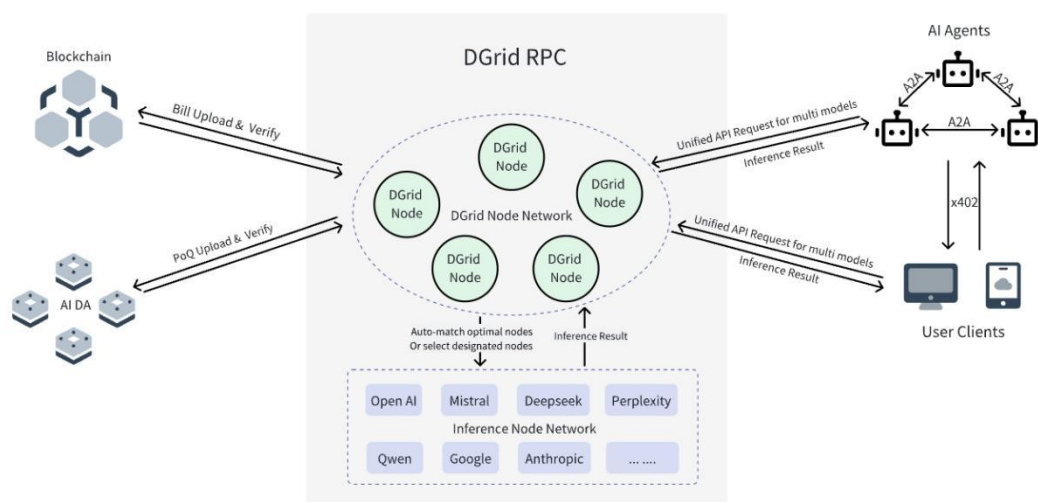
3 Design

DGrid.AI = AI RPC + LLM Reasoning + Distributed Network Nodes

DGrid.AI addresses the critical gaps in Web3 AI and centralized AI limitations

through an interconnected ecosystem of nodes, protocols, and decentralized infrastructure. By integrating standardized AI RPC interfaces, distributed inference nodes, intelligent routing, on-chain settlement, and secure storage, it builds a trustless, scalable, and user-centric LLM inference network—making AI a native capability for blockchain applications. At its core, DGrid's solution redefines decentralized AI inference by integrating three foundational components: distributed nodes for model execution with Proof of Quality (PoQ) to guarantee result credibility, standardized protocols for universal access, and on-chain mechanisms for transparency.

Together, these elements eliminate reliance on centralized providers, enabling AI to operate as an open, community-governed utility.



3.1 DGrid Nodes: Decentralized Inference Execution

DGrid nodes are community-operated nodes that form the computational core of the network by integrating one or more large language models (LLMs, such as Llama-2, Mixtral). These nodes:

- Performs reasoning tasks for users by processing inputs (such as text prompts or smart contract queries) and generating outputs through preloaded models. Simultaneously, it validates the quality of reasoning results via the Proof of Quality (PoQ) mechanism to ensure the reliability and accuracy of the output.
- Adapt to hardware capabilities, with operators selecting models matching their server specs (from lightweight 7B-parameter models on basic GPUs to 70B+ models on high-performance hardware).

- Report metrics in real-time to DGrid adapter nodes (including latency and computational unit [CU] consumption), providing data support for intelligent routing and enabling optimal allocation of tasks.

By distributing inference across thousands of independent nodes, DGrid eliminates single points of failure and ensures geographic redundancy—critical for Web3 applications requiring 24/7 reliability.

3.2 DGridRPC: Universal Access and Request

Verification

- **DGridRPC:** A standardized JSON-RPC protocol simplifying user access to models across the network. It provides a unified API for invoking any LLM (regardless of node or model type) and integrates EIP-712 signatures to validate user requests — ensuring only authorized and prepaid tasks are processed.

DGridRPC resolves the “interface fragmentation” issue in Web3 AI, making LLM integration as straightforward as calling a smart contract.

3.3 PoQ (Proof of Quality): Trustworthy Guarantee for

Inference Results

PoQ (Proof of Quality) is the core mechanism in the DGrid ecosystem that ensures the credibility of LLM inference results. It operates in synergy with distributed nodes and GridRPC to form a closed loop of “request-execution-verification”:

- **Multi-dimensional Quality Assessment:** PoQ conducts objective scoring on inference results generated by DGrid nodes based on three key dimensions: “Accuracy Match” (comparison with standard answers or reference results), “Response Consistency” (output deviation of the same request across different nodes), and “Format Compliance” (adherence to the output format requirements specified in user requests).
- **On-chain Verifiable Proof Generation:** After a node completes an inference task, it must upload the inference process logs and PoQ scoring data to the network to generate tamper-proof quality proofs. Users can query these proofs on-chain to quickly verify the reliability of results without re-executing the inference task.

3.4 Billing Contracts and AI DA Layer: On-Chain Transparency

- **Billing Contract:** A smart contract deployed on the blockchain to automate \$DGA token settlements between users and nodes. This contract calculates fees based on Computation Units (CU) and latency, deducts payments from user accounts via the x402 protocol, and distributes rewards to node operators—eliminating intermediaries.
- **AI DA Layer (Data Availability):** A decentralized storage network where all inference request data is PoQ-backed to ensure auditability. Users can verify billing details, while nodes can prove task completion, enhancing transparency for dispute resolution or compliance audits.

3.5 Security Mechanism

DGrid.AI has established a comprehensive security framework to ensure trustlessness in a decentralized network, combining technical safeguards and on-chain transparency:

3.5.1 Trusted Inference Environments

- **Immutable Runtimes:** DGrid Node operators cannot modify LLM weights or execution environments, ensuring consistent model behavior across the network.
- **Resource Controls:** Strict limits on CPU, GPU, and network usage (enforced by Nodes) prevent denial-of-service attacks.

3.5.2 On-Chain Auditing and Accountability

- **Immutable Records:** All critical activities—node registrations, inference metadata (inputs/outputs), fee settlements, and rewards—are recorded on-chain via the Bill Contract and archived in AI DA.
- **Automated Penalties:** DGrid Nodes monitor node behavior; malicious actors (e.g., submitting fake results) face staked token slashing or jailing, enforced by smart contracts.
- **Decentralized Governance:** \$DGA token holders vote on protocol upgrades, fee structures, and security parameters, ensuring the network evolves in line with community interests.

By combining secure inference environments, on-chain transparency, and community governance, DGrid.AI ensures the network operates in a secure, reliable, and trustless manner—delivering robust decentralized AI inference services to users.

3.6 dToken: Incentives & Governance

\$DGA (dToken) serves as the network's economic engine, aligning interests across the ecosystem:

- **Payments:** Users pay \$DGA for inference tasks, with fees dynamically adjusted via the Bill Contract.
- **Rewards:** Node operators earn \$DGA based on contribution quality (e.g., low latency, high uptime) and validation participation.
- **Staking:** DGrid Nodes must stake \$DGA to participate, with slashing for misconduct.
- **Governance:** Token holders vote on protocol parameters (e.g., fee structures, model whitelisting) to guide network evolution.

This architecture delivers a solution that is scalable (anyone can operate a node), trustless (on-chain proofs replace reliance on intermediaries), and Web3-native (integrated with blockchain workflows). By unifying distributed execution, intelligent coordination, secure reasoning, and transparent settlement, DGrid.AI transforms LLM inference into a foundational capability for Web3—from DeFi strategy analyzers to on-chain chatbots and more.

4 DGrid Node System

GridNodes form the distributed backbone of DGrid.AI, with classifications based on functional specialization, hardware capabilities, and task complexity. Each node type contributes uniquely to the network's resilience, scalability, and trustlessness, collectively eliminating reliance on centralized AI infrastructure. Below is a detailed breakdown of node categories, their roles, and their significance to the decentralized ecosystem.

4.1 Significance of Node Operators to Decentralized AI

Node operators are the foundation of DGrid.AI's mission to democratize AI infrastructure. Their participation delivers three critical benefits:

4.1.1 Eliminating Centralized Control

By distributing node ownership across a global community, DGrid.AI resists censorship, corporate monopolies, or single-entity manipulation. Unlike centralized AI platforms (where a single provider controls access and moderation), no entity can unilaterally restrict, alter, or censor inference results—aligning with Web3's ethos of neutrality.

4.1.2 Enhancing Network Resilience

A geographically diverse operator base ensures the network remains robust against:

- **Regional Outages:** Failures in one region are offset by nodes in others, preventing service disruptions.
- **Regulatory Risks:** Decentralized ownership reduces vulnerability to localized policy changes or bans.
- **Hardware Failures:** Redundancy across thousands of nodes eliminates single points of failure, ensuring 24/7 reliability for Web3 applications.

4.1.3 Driving Ecosystem Growth

Operators contribute more than computing power—they shape the network's evolution:

- **Feedback Loops:** Operators report performance issues (e.g., model bottlenecks) via governance, guiding protocol upgrades.
- **Incentive Alignment:** \$DGA rewards for high performance (low latency, uptime) ensure operators invest in improving node quality, directly benefiting users.
- **Diversity of Offerings:** Operators choose which models to host, fostering a rich ecosystem of LLMs (from open-source to specialized) that cater to diverse Web3 use cases.

In summary, DGrid.AI's node classification ensures specialized roles work in harmony—DGrid Nodes execute inference, nodes coordinate and validate, and operators sustain decentralization. Together, they transform AI from a centralized service into a community-governed utility, accessible to all.

5 Economic System

\$DGAI is the native utility token of the DGrid.AI ecosystem, serving as the backbone of value exchange, incentive alignment, and network governance in our decentralized AI inference network. Designed to operate seamlessly within Web3 infrastructure, it enables trustless interactions between users, node operators, and developers—eliminating reliance on intermediaries.

5.1 Technical Specifications

- **Standard:** BEP-20 token, compatible with BNBChain and all EVM-compatible blockchains, ensuring broad interoperability with wallets, exchanges, and smart contracts.
- **Total Supply:** Fixed at 1 billion tokens, with no inflationary minting post-launch. This scarcity ensures long-term value stability and aligns incentives with network growth.
- **Key Extensions:** Integrates EIP-2612 (permit) for gasless approvals, simplifying payment flows for inference tasks and staking operations.

5.2 Core Functions

\$DGA

I binds the DGrid.AI ecosystem together, fostering a self-sustaining cycle of participation: Users pay \$DGAI for inference, node operators earn \$DGAI for contributions, and staking/governance ensure long-term network health. By embedding \$DGAI into every interaction—from task submission to result verification—we create a decentralized AI economy where value flows directly between contributors and users, aligned with our core philosophy: "AI should serve the open network."

\$DGA

I powers every critical interaction in the DGrid.AI ecosystem, from inference payments to node accountability:

5.2.1 Inference Task Payments

Users pay \$DGA

I to access LLM inference services via DGrid Nodes. Fees are dynamically calculated based on:

- **Compute Units (CU):** A standardized measure of inference complexity (tied to model size, input tokens, and execution time).

- Latency: Adjustments for task speed (rewards for faster execution, minor penalties for delays).

Payments are processed automatically via the Bill Contract, a blockchain-deployed smart contract that deducts \$DGAI from users and distributes funds to node operators—ensuring transparency and immutability.

5.2.2 Node Operator Rewards

\$DGAI incentivizes community participation in network operations:

- DGrid Nodes: Earn \$DGAI based on inference task contributions (e.g., CU processed, low latency, and uptime). Rewards are proportional to performance, with higher payouts for hosting large models (13B+ parameters) or serving high-priority requests.

Rewards are distributed weekly via the Bill Contract, with details archived to AI DA for auditability.

5.2.3 Staking Collateral

To ensure accountability, \$DGAI staking is required for node participation, with tiered requirements based on node type:

- Standard DGrid Nodes
- High-Performance DGrid Nodes

Staked tokens act as collateral against misconduct (e.g., submitting fake inference results, downtime >72 hours). Malicious nodes face slashing (5–20% of staked \$DGAI) or temporary jailing, enforced automatically via smart contracts.

5.2.4 Network Governance

\$DGAI holders govern the evolution of DGrid.AI through on-chain voting, with proposals including:

- Adjusting fee structures (e.g., base CU pricing, latency penalties).
- Whitelisting new LLM models for network support.
- Upgrading core protocols (e.g., GridRPC) or settlement rules.
- Allocating treasury funds for ecosystem growth (e.g., developer grants, security audits).

Voting power is proportional to staked \$DGAI, ensuring decisions align with the interests of active network participants.

5.3 Circulation Mechanism

\$DGAI operates within a closed-loop economic system, ensuring value flows transparently between users, node operators, and the broader ecosystem. This mechanism is governed by smart contracts (e.g., Bill Contract) and enforced by DGrid Nodes, eliminating intermediaries and aligning incentives with network health.

5.3.1 Core Cycle: From Payment to Reward

The circulation of \$DGAI follows a structured flow, linking user demand for inference with node contributions:

1. User Payment Initiation

When a user submits an inference request via a User Client, the process unfolds as:

- The request is routed to a DGrid Node, which calculates the task fee using:
$$\text{Cost} = \text{Compute Units (CU)} \times \text{Base Price} \times (1 + \text{Latency Penalty Factor})$$
 - **CU:** A standardized metric of inference complexity (tied to model size, input tokens, and execution time).
 - **Base Price:** Fixed per CU (adjustable via governance, initially 0.05 \$DGAI).
 - **Latency Penalty Factor:** 0–0.3 (minor adjustments for delays beyond target thresholds).
- The user's \$DGAI is escrowed in the Bill Contract (a blockchain-deployed smart contract) to guarantee payment upon task completion.

2. Reward Distribution to Nodes

Post-inference (verified by Grid Nodes and archived to AI DA), the escrowed \$DGAI is distributed via the Bill Contract. Recipients:

- **DGrid Nodes:** Compensation for executing inference tasks (proportional to CU, latency, and uptime).
- **Treasury:** Funds network maintenance (e.g., AI DA fees, security audits) and developer grants.
- **Governance Pool:** Incentives for \$DGAI holders participating in voting (e.g., protocol upgrades, parameter adjustments).

3. Staking & Lock-Up Requirements

To participate in the network, nodes must lock \$DGAI as collateral, ensuring accountability.

4. Governance & Fee Allocation

- **Governance Pool:** 5% of task payments (plus 10% of Treasury funds annually) are allocated to reward \$DGAI holders who vote on proposals (e.g., adjusting Base Price, whitelisting new models).
- **Network Fees:** A 2% processing fee on all task payments is deducted upfront to cover smart contract gas costs and AI DA archiving, ensuring transaction efficiency.

5. Slashing & Token Burn

Misconduct by nodes triggers automated penalties to protect network integrity:

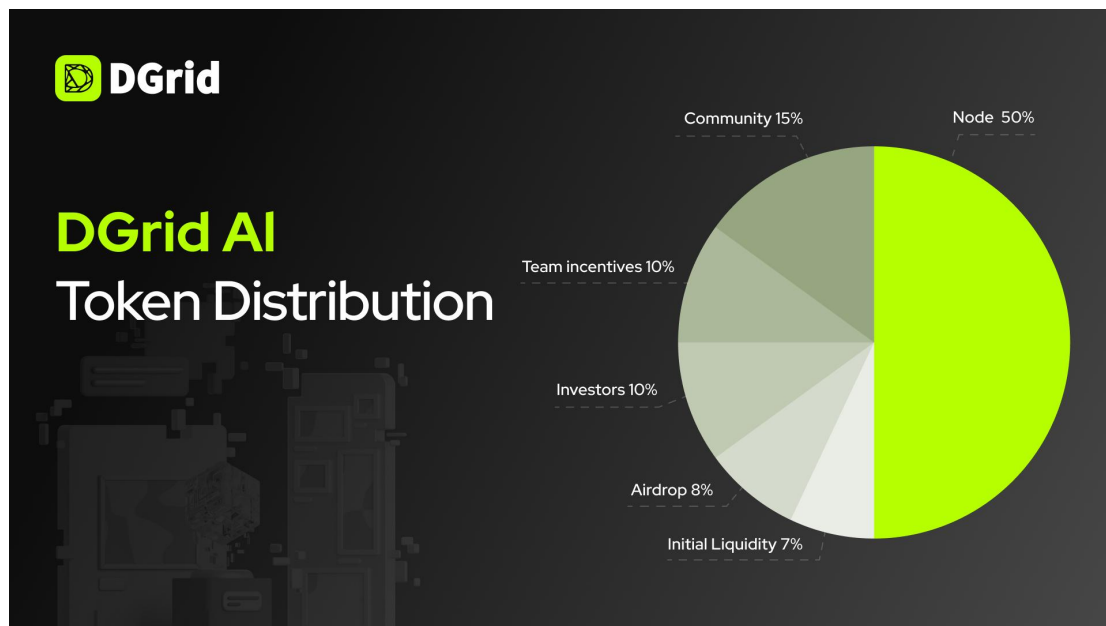
- **Slashed Tokens:** 5–20% of a node's staked \$DGAI is confiscated for malicious behavior (e.g., submitting fake inference results, downtime >72 hours).
- **Burn Mechanism:** All slashed tokens are permanently burned, reducing the total \$DGAI supply over time and enhancing scarcity.

5.3.2 Key Flows: Inflows & Outflows

Inflows (Sources of \$DGAI)	Outflows (Uses of \$DGAI)
User payments for inference tasks	Weekly rewards for DGrid Nodes Nodes
Staking locks for node participation	Governance rewards for voting participants
Network fees (gas, AI DA)	Initial distribution (e.g., public sale, airdrops)
Token buybacks from secondary markets	Slashed tokens (burned)

This mechanism ensures \$DGAI retains utility as both a medium of exchange and a governance tool, fostering a self-sustaining ecosystem where value is directly proportional to network participation. By linking payments to contributions and enforcing accountability via staking, \$DGAI's circulation strengthens the decentralization and resilience of DGrid.AI.

5.4 Token Distribution



\$DGAI has a fixed total supply of 1 billion tokens, with distribution designed to align incentives across users, node operators, developers, and the broader ecosystem. Tokens enter circulation through structured channels, prioritizing long-term network growth over short-term speculation. The 1 billion \$DGAI tokens are distributed as follows:

- **Nodes (50%):** Incentivizes decentralized inference execution and security through DGrid Nodes, adopting a 10-year unlocking mechanism with a halving every 2 years.
- **Community Rewards (15%):** Used to support developer grants, strategic cooperation, cross-chain integration, and community initiatives. 30% is unlocked immediately at TGE (Token Generation Event) for pre-mined token redemption and TGE operation activities; the remaining portion is unlocked through 2-year linear vesting.
- **Team Incentives (10%):** Compensates founding engineers and core contributors to ensure long-term team commitment, using a 2-year linear vesting mechanism.
- **Investors (10%):** Allocated to strategic investors who supported pre-mainnet development. 10% is unlocked immediately at TGE, and the remaining portion is unlocked through 2-year linear vesting.
- **Airdrops (8%):** Incentives related to exchange listings, with 100% unlocked immediately at TGE.
- **Initial Liquidity (7%):** Reserved for market making on centralized exchanges (CEX) and decentralized exchanges (DEX), with 100% unlocked immediately at

TGE.

6 Governance

Governance in DGrid.AI is a decentralized process that empowers \$DGAI token holders—including node operators, users, and developers—to guide the network's evolution. By participating in voting, the community ensures protocol changes align with ecosystem needs, from technical upgrades to economic adjustments.

6.1 Voting Rights: Who Participates?

All \$DGAI token holders can participate in governance, with voting power proportional to their staked tokens. This ensures those invested in the network's success (e.g., node operators, long-term users) have the most influence.

- **Voting Power Calculation:** 1 staked \$DGAI = 1 vote. Unstaked tokens do not contribute to voting power, incentivizing long-term commitment.
- **Node Operator Advantages:** DGrid Nodes and high-performance DGrid Nodes (staking $\geq 50,000$ \$DGAI) receive a 10% voting power boost, recognizing their critical role in network stability.

6.2 What Can Be Voted On?

The community votes on key decisions that shape DGrid.AI's functionality, security, and economics:

6.2.1 Protocol Upgrades

- Changes to GridRPC (e.g., adding new API endpoints for multi-modal models).
- Updates to Grid Nodes (e.g., enhancing container isolation for larger models).
- Integrations with new blockchains or AI DA networks.

6.2.2 Economic Parameters

- Adjustments to the Base Price per Compute Unit (CU).
- Latency penalty factors (e.g., reducing penalties for regions with infrastructure limitations).
- Reward splits from DGrid Nodes.

6.2.3 Model Governance

- Whitelisting new LLMs for network support (e.g., approving a 70B open-source model).
- Blacklisting models with security risks or poor performance.

6.2.4 Treasury Allocation

- Funding for developer grants (e.g., building GridRPC SDKs for Solidity).
- Allocations for security audits or bug bounties.

6.3 The Governance Process

Proposals follow a structured, on-chain workflow to ensure transparency and efficiency:

6.3.1 Proposal Submission

- Any \$DGAI holder can submit a proposal, provided they:
 - Deposit a 5,000 \$DGAI "bond" (refunded if the proposal reaches quorum; forfeited otherwise to prevent spam).
 - Include technical specifications, cost-benefit analysis, and implementation timelines.
- Proposals are hosted on-chain and mirrored to AI DA for public review.

6.3.2 Discussion Period (7 Days)

- The community debates the proposal via official channels (Telegram, forum), with node operators and developers providing technical feedback.
- The proposer may revise the proposal based on feedback, with changes logged on-chain.

6.3.3 Voting Period (7 Days)

- Eligible voters (staked \$DGAI holders) cast votes: For, Against, or Abstain.
- Quorum Requirement: At least 30% of staked \$DGAI must participate for the vote to be valid.

- **Passing Threshold:** Proposals require $\geq 51\%$ "For" votes to be approved.

6.3.4 Execution

- **Approved Proposals:** Automatically implemented via smart contract upgrades (for on-chain parameters) or network-wide node updates (for protocol changes). Nodes enforce compliance, penalizing non-upgraded nodes.
- **Rejected Proposals:** The 5,000 \$DGAI bond is transferred to the Treasury to fund ecosystem growth.

6.4 Why Decentralized Governance Matters

- **Alignment with Community Needs:** Decisions are made by those using and operating the network, not centralized teams.
- **Adaptability:** The network can evolve quickly to adopt new technologies (e.g., integrating a breakthrough LLM) or address challenges (e.g., adjusting fees during high demand).
- **Trustless:** On-chain voting eliminates reliance on intermediaries, ensuring outcomes are enforced by code, not human discretion.

By empowering \$DGAI holders to govern key decisions, DGrid.AI ensures the network remains community-driven, innovative, and aligned with its core mission: making AI a decentralized utility for Web3.

7 The Future

In the future empowered by DGrid, large language model (LLM) inference will no longer be a privilege hoarded by centralized platforms; instead, it will evolve into a decentralized utility accessible to every participant in the open network. This new era of AI will be defined by transparency, verifiability, and community ownership—where AI's power is equitably shared, and trust is enforced not by intermediaries, but by code and collective participation.

For AI enthusiasts and developers, DGrid will become an innovative playground unshackled by centralized gatekeepers. They will log into the network to deploy custom LLMs on DGrid Nodes, fine-tune models using open or proprietary datasets, and share their creations via a unified interface. With secure, isolated execution environments, developers retain full control over their AI assets—earning rewards for contributions while ensuring their work remains transparent and auditable.

Data scientists will thrive in DGrid's open ecosystem, leveraging the network's vast

distributed pool of models and computing resources to accelerate research breakthroughs. From optimizing LLM efficiency for edge devices to building specialized Web3 tools (such as real-time DeFi risk analyzers and smart contract auditors), their work will drive tangible progress across AI applications. For business leaders, DGrid will redefine enterprise AI integration—enabling cost-efficient, customizable AI solutions without vendor lock-in or data silos.

Even individuals without deep AI expertise will find meaningful roles in DGrid's ecosystem. They will operate DGrid Nodes with minimal technical expertise—turning idle GPU resources into active network contributors and earning rewards for their participation. They will also engage in governance, voting on model prioritization, fee structures, and ecosystem upgrades to ensure the network evolves in alignment with community needs.

At its core, the DGrid ecosystem is more than a technical network; it is a foundational movement to democratize AI infrastructure. It transforms LLM inference from a centralized service into a community-governed utility, where every participant—whether developer, node operator, enterprise, or casual user—holds a stake in shaping its future. In this vision, AI serves the open network: transparent, verifiable, and empowering, guiding Web3 and beyond toward a future where AI's benefits are shared by all.