



DGrid AI

The Decentralized AI Smart Network

Litepaper - June 2025

<https://dgrid.ai>

1 Abstract

With the core mission of "building a trusted, efficient AI network connecting supply and demand", DGrid addresses the industry pain points of market fragmentation and centralized control. We break down the fragmentation and value mismatch in the AI ecosystem, providing a one-stop solution for both Web3 and traditional AI sectors. Adhering to the core principle that "AI should serve the open web", we are committed to enabling high-quality AI capabilities (including LLMs and AI Agents) to achieve efficient flow, fair distribution, and trusted verification through technological innovation and ecosystem collaboration.

To tackle the core challenges—such as the explosion of model supply alongside fragmented interfaces, high integration costs for developers, value monopolization of creators, and non-auditable inference results—DGrid has built a three-tier architecture: "Decentralized Routing & Verification Network + LLM & Agent Free Market + AI DAO", which connects the entire chain from AI capability supply to demand matching:

- Through a unified API and intelligent routing, we resolve interface fragmentation. Developers can access thousands of models and Agents without redundant development, significantly lowering integration barriers.
- We build a trusted inference environment using the Proof of Quality (PoQ) algorithm, enabling verifiability and traceability of AI execution processes and output results, and bridging the trust gap in Web3 AI.
- We establish an open LLM & Agent free market, allowing creators such as model providers and prompt engineers to set their own prices and profit directly from global users, breaking platform monopolies.
- Via a distributed node network and elastic scheduling mechanism, we address the high costs, single points of failure, and censorship risks of centralized AI, while providing a merit-based incentive system for node operators.

Ultimately, DGrid aims to become the "value hub" of the AI and Web3 ecosystems. We not only deliver cost-effective, verifiable, high-quality AI services to users but also build a fair and win-win collaboration platform for ecosystem participants, advancing AI from "centralized services" to a "foundational public utility of the open web".

Core Values

- **Decentralization:** AI belongs to the open web, not centralized platforms, ensuring equal participation of all ecosystem stakeholders.
- **Accessibility:** Anyone can join as a node operator, developer, creator, or user, with low barriers to entry.

- **Trustlessness:** Based on on-chain permission and auditing mechanisms, services are verifiable and traceable without relying on third-party trust.
- **Scalability:** Modular design enables flexible model deployment, inference, and caching, adapting to diverse scenario requirements.
- **Value Discoverability:** Intelligent routing makes inference resources "routable, verifiable, and comparable in price", enabling efficient allocation of intelligent liquidity.

2 Background

In the current field of artificial intelligence, both centralized systems and Web3 ecosystems face key constraints that hinder the seamless integration of artificial intelligence and blockchain. The AI industry is experiencing explosive growth, yet it lacks both market efficiency and fairness, which restricts the overall development of the ecosystem.

Market Pain Points:

- **Explosive Supply and Fragmented Ecosystem:** The number of proprietary, open-source, fine-tuned models and AI Agents is surging. However, inconsistent interface specifications and drastic quality disparities make it difficult for resources to interoperate and be compatible with each other.
- **Matching Failure and Value Obscuration:** For developers, the cost of finding, testing, and integrating suitable models and Agents is extremely high, leading to low efficiency. For creators, excellent models and Agents lack efficient distribution channels, with their value monopolized by centralized platforms, resulting in an imbalanced profit distribution.
- **Lack of Trust and Concentration of Power:** The service process operates in a "black box" mode—AI output results are neither verifiable nor auditable, and there is no room for autonomous governance. Centralized hierarchies control pricing power and traffic distribution, depriving the ecosystem of transparency and fairness.

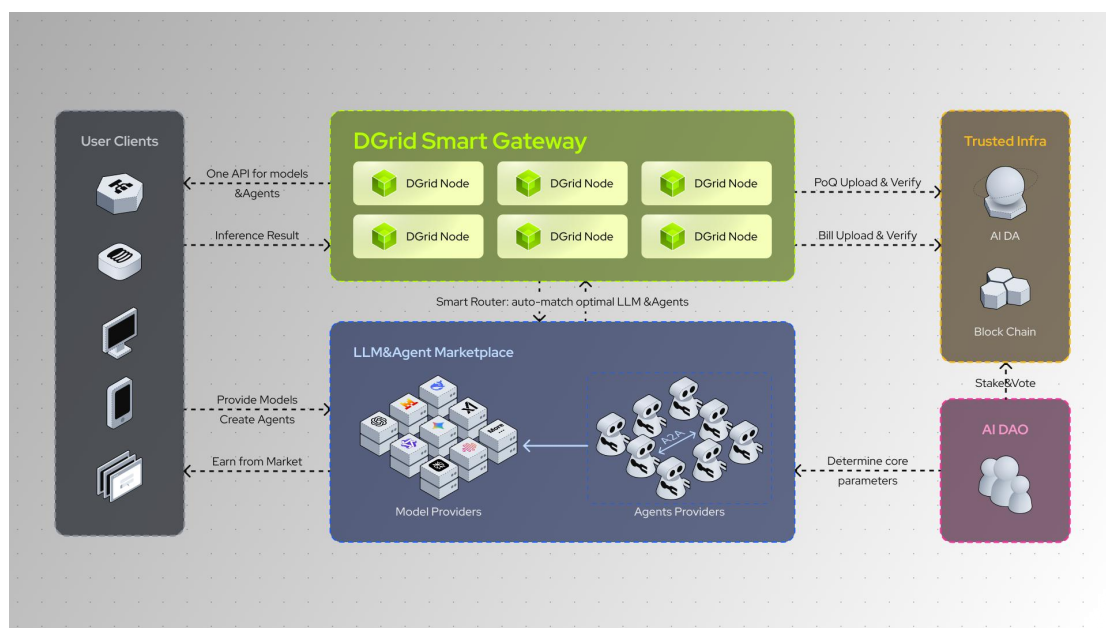
The mission of DGrid.AI is to "reconstruct the underlying architecture of AI inference" by building decentralized infrastructure, freeing AI operations and applications from dependence on a few large platforms. The ultimate goal is to make artificial intelligence a native foundational capability of the blockchain world—seamlessly integrated into blockchain applications, just like current data storage and transaction processing.

3 Design

DGrid.AI = Decentralized Routing & Verification Network + LLM & Agent Free Market + AI DAO Governance System

DGrid.AI addresses the critical gaps between Web3 AI and centralized AI through an interconnected ecosystem composed of nodes, protocols, and decentralized infrastructure.

By integrating standardized AI RPC interfaces, distributed inference nodes, intelligent routing, on-chain settlement, and secure storage, we have built a trustless, scalable, and user-centric LLM and Agent service network—enabling AI to become a native capability of blockchain applications. At its core, DGrid’s solution redefines decentralized AI inference through three foundational components: distributed nodes for model execution with result trustworthiness guaranteed by Proof of Quality (PoQ), standardized protocols for universal access, and on-chain mechanisms for full transparency.



3.1 Decentralized Routing & Verification Network

- **Unified Intelligent Access:** Provides developers with a single API to access thousands of specialized models and preconfigured Agents simultaneously, significantly reducing integration costs.
- **Intelligent Routing & Scheduling:** Automatically recommends and schedules optimal service resources based on multi-dimensional metrics including task type,

budget constraints, and historical performance data.

- **Trusted Verification Mechanism:** Leverages the Proof of Quality (PoQ) algorithm to ensure the trustworthiness of AI and Agent workflows and output results, achieving efficient, censorship-resistant, and traceable services.

3.2 LLM & Agent Free Market

- **Open Listing Opportunities:** Model providers, AI developers, and prompt engineers can list fine-tuned models or customized Agents on the DGrid network, set their own prices, and earn revenue directly from global users through a transparent competition mechanism.
- **Value Tokenization:** High-quality models and Agents can be tokenized on DGrid to capture long-term market value.

3.3 DGrid Nodes: Decentralized Inference Execution

DGrid Nodes are community-operated nodes that form the computational core of the network by hosting one or more Large Language Models (LLMs, e.g., Llama-2, Mixtral) and AI Agents. These nodes:

- Execute inference tasks for users, process inputs (e.g., text prompts, smart contract queries), generate outputs via preloaded models, and verify inference result quality through the PoQ (Proof of Quality) mechanism to ensure output trustworthiness and accuracy.
- Adapt to hardware capabilities: Operators can select models matching their server specifications—ranging from lightweight 7B-parameter models on basic GPUs to 70B+-parameter models on high-performance hardware.
- Report real-time metrics (latency, Compute Unit [CU] consumption) to the DGrid network, providing data support for intelligent routing and enabling optimized task allocation.

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

3.4 DGridRPC: Universal Access&Request Verification

DGridRPC: A standardized JSON-RPC protocol that simplifies user access to models and Agents in the network. It provides a unified API for invoking any LLM or

Agent (regardless of node or model type) and integrates EIP-712 signatures to verify user requests—ensuring only authorized and pre-paid tasks are processed.

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

3.5 Proof of Quality (PoQ): Trust Guarantee for Inference Results

Proof of Quality (PoQ) is the core mechanism in the DGrid ecosystem that ensures the trustworthiness of LLM inference and Agent execution results. Working in tandem with distributed nodes and DGridRPC, it forms a closed-loop "Request-Execution-Verification" workflow:

- **Multi-Dimensional Quality Assessment:** PoQ objectively scores inference results generated by DGrid Nodes based on three key dimensions: accuracy alignment (comparison against standard answers or reference results), response consistency (output deviation of the same request across different nodes), and format compliance (adherence to user-specified output requirements). The evaluation framework further incorporates cost-efficiency considerations and semantic similarity verification logic to balance high quality with low overhead.
- **On-Chain Verifiable Proof Generation:** After completing an inference task, nodes upload inference process logs and PoQ score data to the network to generate tamper-proof quality proofs. Users can query these proofs on-chain to quickly verify result reliability without re-executing inference tasks. Combined with a hybrid verification architecture, this mechanism achieves low latency and low cost while ensuring computational integrity.

3.6 Billing Contracts & AI DA Layer: On-Chain Transparency

- **Billing Contracts:** Smart contracts deployed on the blockchain for automated \$DGAI token settlement between users and nodes. These contracts calculate fees based on Compute Units (CU) and latency, deduct payments from user accounts via the x402 protocol, and distribute rewards to node operators—eliminating intermediaries.

- **AI DA Layer (Data Availability):** A decentralized storage network where all inference request data is secured by PoQ to ensure auditability. Users can verify billing details, and nodes can prove task completion, thereby increasing transparency for dispute resolution or compliance audits.

3.7 Security Mechanisms

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

3.7.1 Trusted Inference Environment

- **Immutable Runtime:** DGrid Node operators cannot modify the weights or execution environments of Large Language Models (LLMs) and AI Agents, ensuring consistent model behavior across the network.
- **Resource Controls:** Strict limits on CPU, GPU, and network usage (enforced by nodes) prevent Denial of Service (DoS) attacks.

3.7.2 On-Chain Auditing & Accountability

- **Tamper-Proof Records:** All critical activities—node registration, inference metadata (inputs/outputs), fee settlements, and rewards—are recorded on-chain via Billing Contracts and stored in the AI DA Layer.
- **Automatic Penalty Mechanisms:** The DGrid network monitors node behavior; malicious actors (e.g., submitting false results) face penalties such as confiscation of staked tokens or node blacklisting, enforced by smart contracts.
- **Decentralized Governance:** \$DGAI token holders can vote on protocol upgrades, fee structures, and security parameters, ensuring network development aligns with community interests.

3.8 Core Advantages: Defining the Next-Generation AI Infrastructure

By integrating decentralized execution, intelligent coordination, secure inference, and transparent settlement, DGrid builds core advantages that distinguish it from traditional centralized AI and existing Web3 AI solutions—redefining the next generation of AI infrastructure:

- **End-to-End Trustworthiness:** Based on PoQ quality proofs and on-chain evidence storage, it achieves verifiable and auditable inference processes and results, solving the "black box" problem.
- **Seamless Integration Experience:** A single DGridRPC interface is compatible with thousands of models and Agents, significantly reducing developer integration and migration costs.
- **Ecosystem Fairness & Win-Win Collaboration:** The open LLM & Agent Free Market and merit-based incentive mechanism ensure value alignment among creators, builders, and users.
- **Elasticity & Resilience:** The distributed node network eliminates single points of failure, and geographic redundancy guarantees 24/7 service availability, resisting regional outages and regulatory risks.
- **Community-Driven Autonomous Governance:** AI DAO drives network evolution; \$DGAI holders lead key decisions, and AI Agents can participate in automated governance on behalf of users, ensuring ecosystem development aligns with community interests.

4 DGrid Node System

Node Operators are the cornerstone of DGrid.ai's mission to democratize AI infrastructure. Their participation delivers three critical advantages.

4.1 Eliminating Centralized Control

By decentralizing node ownership across a global community, DGrid.ai resists censorship, corporate monopolization, or manipulation by any single entity. Unlike centralized AI platforms—where a sole provider controls access and moderation—no entity can unilaterally restrict, modify, or censor inference results, aligning with the ethos of Web3 neutrality.

4.2 Enhancing Network Resilience

A geographically diverse operator base ensures the network can withstand the following risks:

- **Regional Outages:** Failures in one region are offset by nodes in other regions, 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 reliable operation for Web3 applications.

4.3 Driving Ecosystem Development

Operators' contributions extend beyond computational power—they shape the network's evolutionary trajectory:

- **Feedback Loop:** Through governance mechanisms, operators report performance issues (e.g., model bottlenecks) to guide protocol upgrades.
- **Incentive Alignment:** \$DGAI rewards high performance (low latency, high availability), incentivizing operators to invest in node quality improvements that directly benefit users.
- **Service Diversity:** Operators can independently choose the models and Agents to host, fostering a diverse ecosystem of LLMs and Agents (ranging from open-source to specialized) to meet the varied needs of Web3 application scenarios.

In summary, DGrid.ai's node operator ecosystem ensures specialized roles work in synergy—DGrid Nodes execute inference, the network coordinates and verifies, and operators uphold decentralization. Together, they transform AI from a centralized service into a community-governed public 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:** BRC-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

As the native token of the network, \$DGAI undertakes four core functions—staking, payment, incentives, and governance—constructing a positive ecological cycle and driving every key interaction within the DGrid.AI ecosystem.

5.2.1 Staking and Access

Nodes and AI service providers must stake \$DGAI to gain network traffic, establish a reputation system, and ensure network security. Staking requirements are tiered based on node type (node performance), and staked tokens serve as collateral for violating behaviors (e.g., submitting false inference results, downtime exceeding 72 hours). Malicious nodes will face confiscation of 5 – 20% of staked \$DGAI or temporary deactivation; these measures are automatically enforced through smart contracts.

5.2.2 Payment for Inference Tasks

Users pay with \$DGAI via DGrid nodes to access Large Language Model (LLM) inference and Agent services. Fees are dynamically calculated based on the following factors:

- Compute Unit (CU): A standardized metric for measuring inference complexity (linked to model size, number of input tokens, and execution time).
- Latency: Adjusted based on task speed (faster execution earns rewards, while latency incurs minor penalties).

Payments are automatically processed through the Bill Contract, a smart contract deployed on the blockchain. This contract deducts \$DGAI from user accounts and distributes funds to node operators, service providers, etc., ensuring transparency and immutability of transactions.

5.2.3 Rewards for Node Operators

The network rewards high-quality nodes and AI service providers with \$DGAI based on service quality, usage volume, and user evaluations to stimulate enthusiasm for contributions. Operators earn \$DGAI based on their contributions to inference tasks (e.g., number of CUs processed, low latency, and uptime). Rewards are proportional to performance—nodes hosting large models (over 13B parameters) or processing high-priority requests receive higher rewards. Rewards are distributed weekly via the Bill Contract, with detailed records archived in the AI DA Layer to ensure auditability.

5.2.4 Network Governance

\$DGAI is a governance token; token holders have the right to vote on key matters related to network development, ensuring decisions align with the interests of active network participants. Proposals include:

- Adjusting the fee structure (e.g., base CU pricing, latency penalties).
- Adding new Large Language Models (LLMs) to the whitelist for network support.
- Upgrading core protocols (e.g., DGridRPC) or settlement rules.
- Allocating treasury funds to promote 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 process that links users' inference demands to node contributions:

1. Initiation of User Payment

When a user submits an inference request via the user client, the process unfolds as follows:

- The request is routed to a DGrid Node, which calculates the task fee using the following formula: $\text{Cost} = \text{Compute Unit (CU)} \times \text{Base Price} \times (1 + \text{Latency Penalty Coefficient})$
 - **Compute Unit (CU):** A standardized metric for measuring inference complexity (correlated with model size, number of input tokens, and execution time).
 - **Base Price:** A fixed price per CU (adjustable via governance mechanisms, with an initial value of 0.05 \$DGAI).
 - **Latency Penalty Coefficient:** Ranges from 0 to 0.3 (fine-tuned for latency

exceeding the target threshold).

- The user's \$DGAI is escrowed via the Bill Contract—a smart contract deployed on the blockchain—to ensure payment upon task completion.

2. Distribution of Node Rewards

After inference is completed (verified by Grid Nodes and archived in the AI DA Layer), the escrowed \$DGAI is distributed via the Bill Contract. Recipients include:

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

3. Staking and Lock-Up

Requirements Nodes participating in the network must lock up \$DGAI as collateral to ensure accountability.

4. Governance and Fee Allocation

- **Governance Pool:** 5% of task payments (plus 10% of annual treasury funds) are distributed to \$DGAI holders who participate in proposal voting (e.g., adjusting base prices, whitelisting new models).
- **Network Fees:** A 2% processing fee is deducted upfront from all task payments to cover smart contract gas costs and AI DA Layer archiving, ensuring transaction efficiency.

5. Penalty Mechanisms and Token Burning

Node violations trigger automatic penalties to protect network integrity:

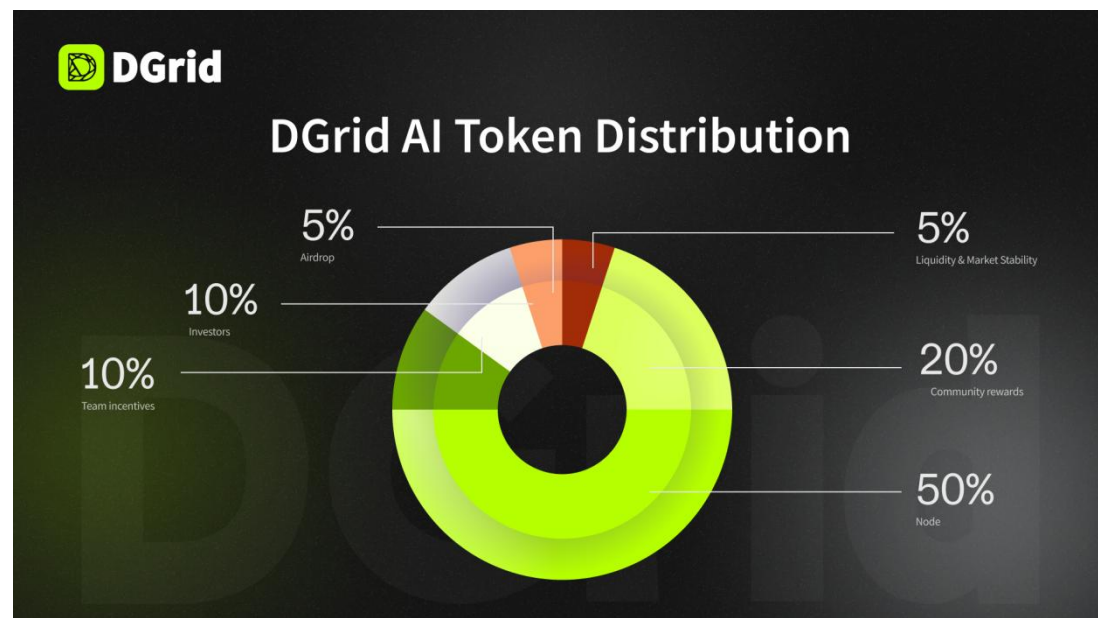
- **Confiscated Tokens:** 5–20% of a node's staked \$DGAI will be confiscated for malicious behaviors (e.g., submitting false inference results, downtime exceeding 72 hours).
- **Burning Mechanism:** All confiscated tokens are permanently burned, reducing the total supply of \$DGAI over time and enhancing scarcity.

5.3.2 Key Flows: Inflows & Outflows

Inflows (Sources of \$DGAI)	Outflows (Uses of \$DGAI)
Fees paid by users for inference tasks	Rewards for node operators for inference tasks
Weekly rewards for DGrid Nodes	Staking lock-up for node participation
Governance rewards for voting participants	Network fees (gas, AI DA Layer archiving)
Initial distribution (e.g., public sale, airdrops)	Confiscated tokens (burned)

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

5.4 Token Distribution



The total supply of \$DGAI is fixed at 1 billion tokens. Its distribution design aims to align incentive mechanisms among users, node operators, developers, and the broader ecosystem. Tokens enter circulation through structured channels, prioritizing

long-term network growth over short-term speculation. The distribution of 1 billion \$DGA tokens is as follows:

- **Nodes (50%):** By incentivizing Genesis Premium members, Smart Gateway operators, and LLM Providers, we maintain the security and stability of the decentralized inference network. This mechanism adopts a 10-year mining cycle, with mining reward halving events every two years.
- **Community Rewards (20%):** Used to drive developer grants, strategic collaborations, cross-chain integrations, and community initiatives. 30% is unlocked at TGE (Token Generation Event) for pre-mined token redemption and TGE operational activities; the remainder is subject to 2-year linear vesting.
- **Team Incentives (10%):** Compensates founding engineers and core contributors to ensure long-term team commitment, with a 2-year linear vesting mechanism.
- **Investors (10%):** Allocated to strategic investors supporting pre-mainnet development. 10% is unlocked at TGE, and the remainder undergoes 2-year linear vesting.
- **Airdrops (5%):** Incentives related to exchange listings, with 100% unlocked at TGE.
- **Initial Liquidity (5%):** Reserved for market making on Centralized Exchanges (CEX) and Decentralized Exchanges (DEX), with 100% unlocked at TGE.

6 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.