DLResearch × 0G

Building the Foundation for an Open AI Economy:

The Case for 0G

Building the Foundation for an Open AI Economy: The Case for 0G

Executive Summary

The AI Infrastructure Challenge

Artificial intelligence has become the defining compute market of the next decade. Transformer architectures and scaling laws turned progress into an infrastructure race, not a search for a single breakthrough algorithm. Training cycles now consume millions of GPU hours, while inference has shifted demand from periodic spikes to an always-on tide of queries embedded in daily workflows.

Even as the cost per query has fallen, usage has compounded faster, pushing aggregate spending higher and stressing the entire supply chain from silicon and packaging to networking, storage, and power. Hyperscalers have responded with unprecedented capital programs, but the result for builders is a landscape that is powerful yet closed, efficient yet opaque, and optimised for central platforms rather than open markets.

0G addresses this gap by treating AI infrastructure as a coherent operating system rather than a stack of disconnected services. It brings storage, data availability, compute, and consensus into one modular environment that can scale horizontally and verify outcomes at every step.

The goal is not only to lower costs or increase throughput, but to make AI infrastructure auditable, programmable, and open to permissionless participation. That combination is what turns infrastructure from a black box into a public good.

Data Architecture and Storage

The architecture starts with data. 0G Storage separates immutable archives from fast, mutable state so that the same system can serve training corpora and live application backends without trade-offs. Proof of Random Access requires providers to retrieve specific chunks quickly and reliably, so rewards are paid for useful I/O, not just raw capacity.

This aligns incentives with what AI workloads need in practice: persistence with provenance for large files, and low-latency reads and writes for agent memory, indices, and application state.

Data Availability and Verification

On top of storage, 0G builds a data availability layer that treats publication and persistence as one continuous guarantee. Data is erasure-coded and sampled by quorums selected with verifiable randomness, then finalised by validators that anchor security at a root layer.

Because availability checks operate against data that is actually stored, retrieval is as trustworthy as publication. This matters for AI because models do not only post data once. They fetch, update, and verify repeatedly across training, evaluation, and deployment.

Privacy-Preserving Ownership (ERC-7857)

Beyond the core layers, 0G extends trust to the ownership of digital and AI assets. ERC-7857 introduces on-chain privacy and verifiability by embedding encrypted metadata and secure transfer logic directly into the token standard.

It enables AI models and datasets to be exchanged securely, preserving both confidentiality and authenticity through trusted execution and zero-knowledge proofs. Unlike frameworks such as x402, ERC-8004, Virtual ACP, Google A2A, or Stripe ACP, which focus on payments or identity, ERC-7857 protects the data layer itself. Integrated with 0G Storage and Data Availability, it provides sub-second, encrypted asset transfer for secure and scalable AI ownership.

Compute Marketplace

Compute is organised as an open marketplace. Developers fund workloads for inference, fine-tuning, and eventually training. Providers register GPU capacity and receive jobs that settle through smart contract escrow once proofs confirm correct execution.

Trusted execution and zero-knowledge techniques allow verification without exposing inputs. The marketplace aggregates everything from enterprise clusters to independent operators and pays for validated work rather than promises. This widens supply, reduces lock-in, and lets applications scale without a single vendor gate.

Consensus and Security

Consensus ties the system together. Rather than forcing every function through a single chain, 0G runs many parallel networks that share security through a common stake anchored to Ethereum.

Misbehaviour anywhere is slashable at the root, so guarantees are uniform across storage, availability, and compute. CometBFT provides deterministic finality with sub-second latencies, and the roadmap moves toward parallel confirmation to match the micro-transaction patterns of agent ecosystems. The result is a coordination fabric that grows horizontally while keeping one trust model.

Position in the Market

This design positions 0G clearly within the current market. General-purpose chains deliver either high throughput or deep liquidity, but they externalise the heaviest data flows. Rollups increase capacity, but each one must assemble its own availability and storage, which fragments guarantees.

Dedicated DA layers confirm publication efficiently, yet they stop short of persistence and mutable state. Storage protocols excel at permanence or addressing, but they do not natively provide fast updates or verifiable retrieval tied to a global security model. 0G consolidates these functions in one place and aligns them with the realities of AI workloads.

For developers, the practical advantages are straightforward. Training datasets can be anchored immutably in the same system that serves low-latency key-value access for live applications. Availability is not a temporary property of blobs but a durable property of stored data.

Compute is procured from a global pool and paid only when outputs verify. Coordination is fast enough to support agent networks that read, write, and transact continuously. The net effect is lower integration risk, clearer performance envelopes, and a simpler path from prototype to production.

For users and enterprises, the benefits are clarity and control. Provenance is observable rather than implied. Data can be segmented by policy and location while still participating in a shared marketplace. Pricing can clear in real time, and rights can be enforced with decentralised identity and programmable governance. This is how an open AI economy gains the reliability required for adoption in regulated and mission-critical contexts.

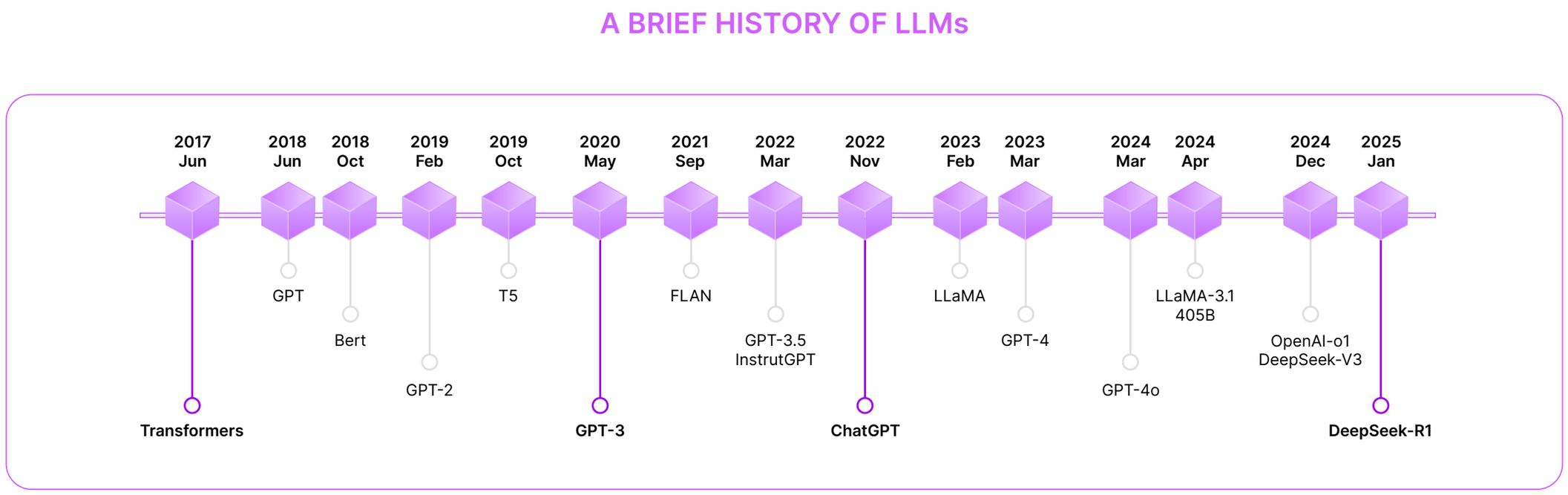
Taken together, these properties show why 0G is more than just another infrastructure option. It offers a unified foundation where data, compute, and availability work as one system. In doing so, it reduces the compromises developers face, provides the assurances enterprises demand, and ultimately sets the standard for decentralised AI infrastructure.

Market Opportunity: AI x Crypto

The AI Market

The Technical Breakthrough

The current wave of AI is powered by a combination of technical breakthroughs and massive capital inflows. In the late 2010s, researchers introduced transformer architectures and discovered scaling laws showing that model performance improves predictably with more parameters, larger datasets, and greater compute. This insight changed the nature of progress in AI: instead of relying on novel algorithms, advancement became a question of who could marshal the most infrastructure.

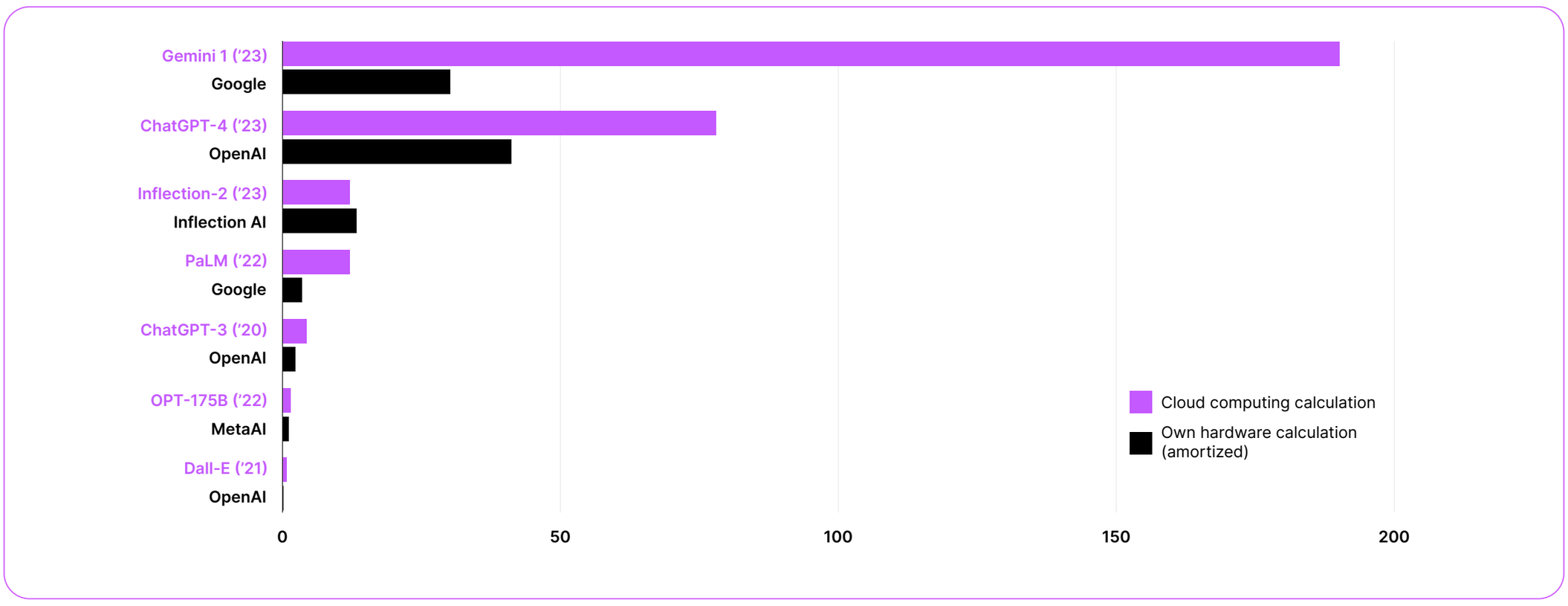


Source: Medium

The pattern since then has been clear. Each new frontier model has been larger, trained on more data, and delivered stronger capabilities. But with every leap in performance, the resource requirements have expanded just as dramatically. Training runs now consume millions of GPU hours, with total compute needs rising four to five times per year through 2024.

Once these systems left the lab and entered products, the pressure on infrastructure only intensified. Training occurs in large but occasional cycles, while inference, the act of serving outputs to users, is relentless.With tools like ChatGPT, Claude, and Gemini embedded into consumer apps and enterprise workflows, every email drafted or line of code generated becomes another inference call. Billions of these queries now take place daily, translating into a constant, compounding demand for compute.

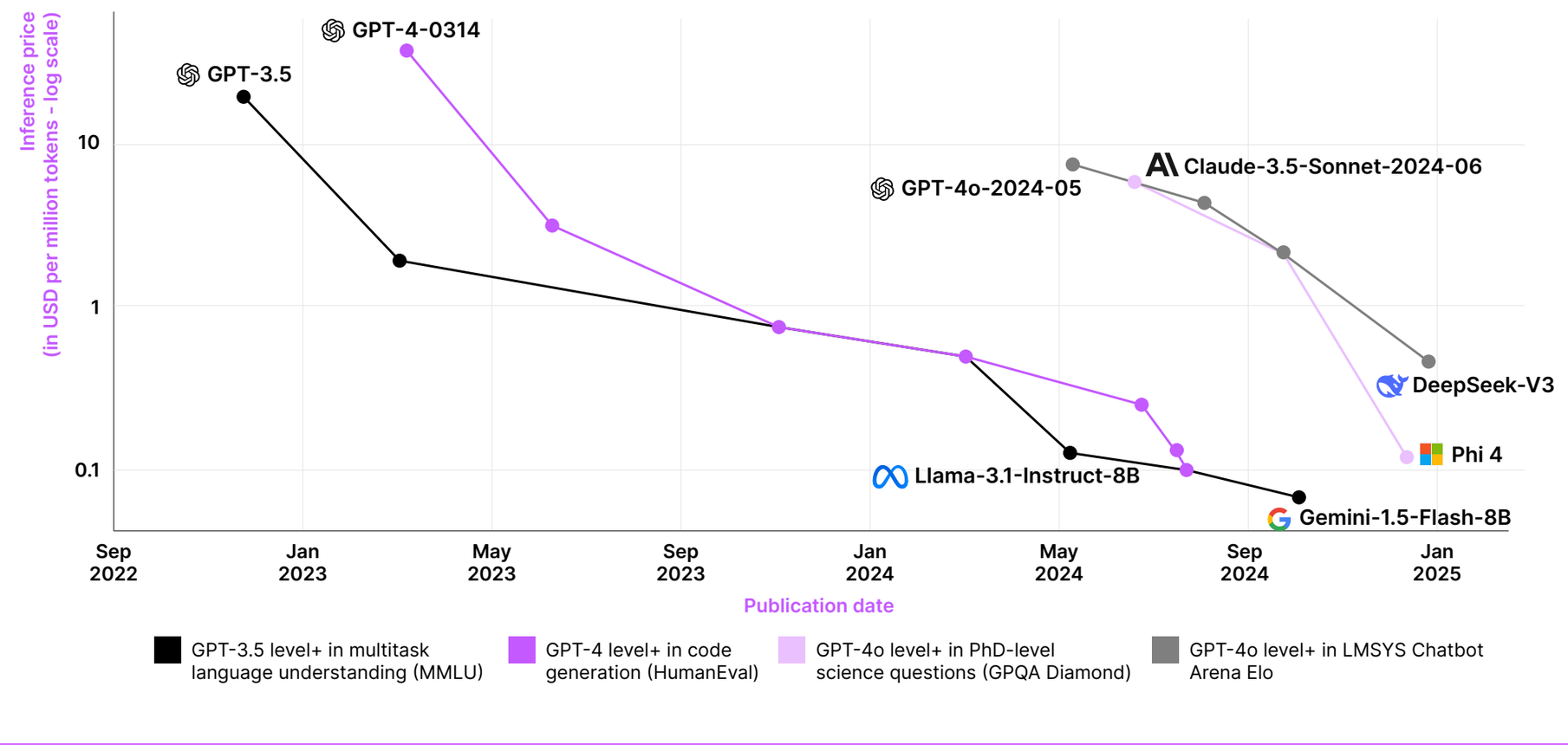
THE EXTREME COST OF TRAINING AI MODELS



Source: Statista

Although the cost per inference has fallen dramatically, with Stanford's AI Index reporting a more than 280-fold drop in the price of GPT-3.5-class queries between late 2022 and late 2024, these efficiency gains have been overwhelmed by rising demand.

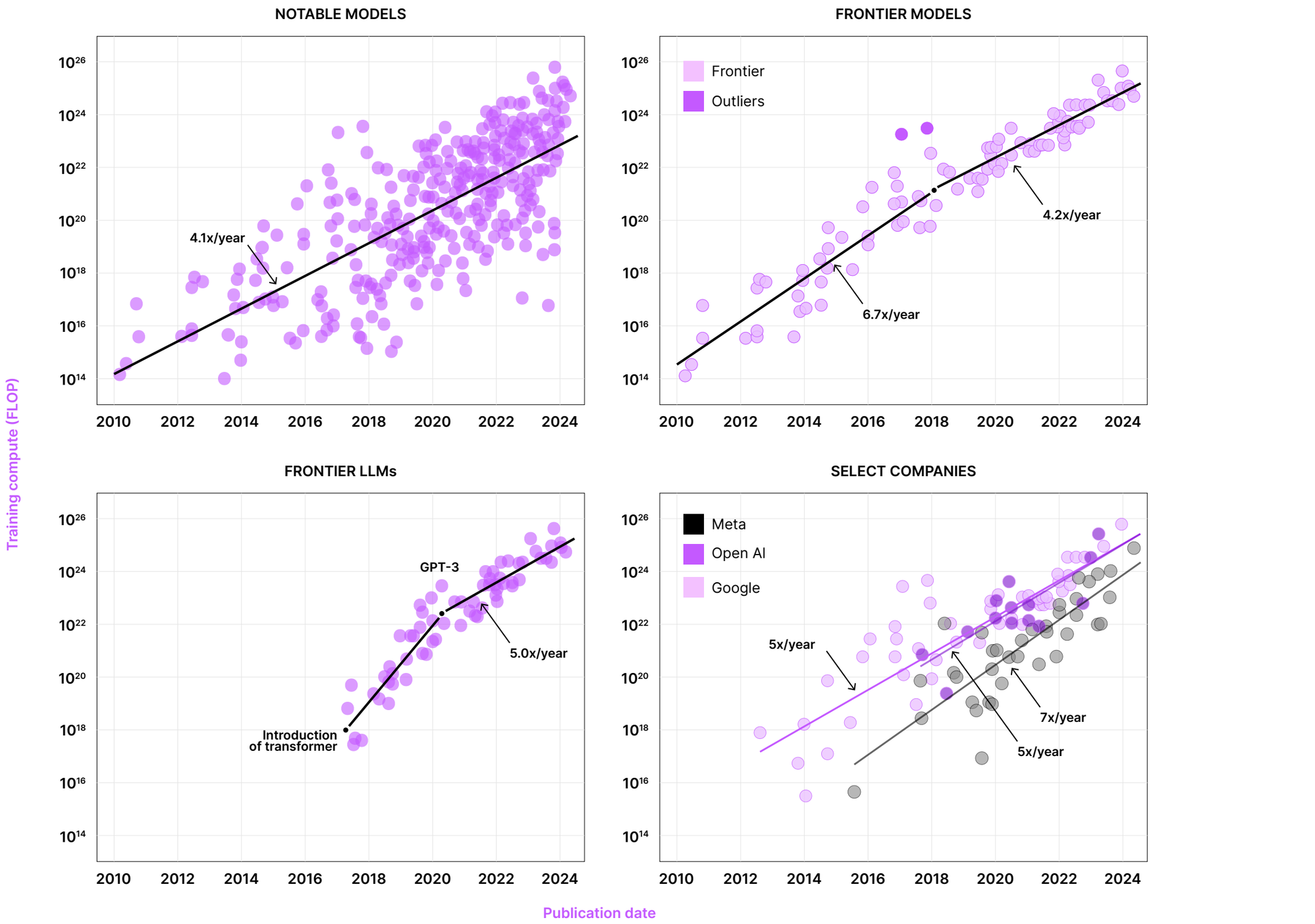
INFERENCE PRICE ACROSS SELECT BENCHMARKS 2022-2024



Source: Rdworldonline

ChatGPT alone now handles an estimated 2.5 billion prompts every day from over 700 million weekly users, a scale that pushes inference workloads far beyond what earlier models faced. At the same time, the cost of training frontier models continues to climb, growing by roughly 2.4 times per year since 2016.

SUMMARY OF COMPUTE TRENDS IN AI



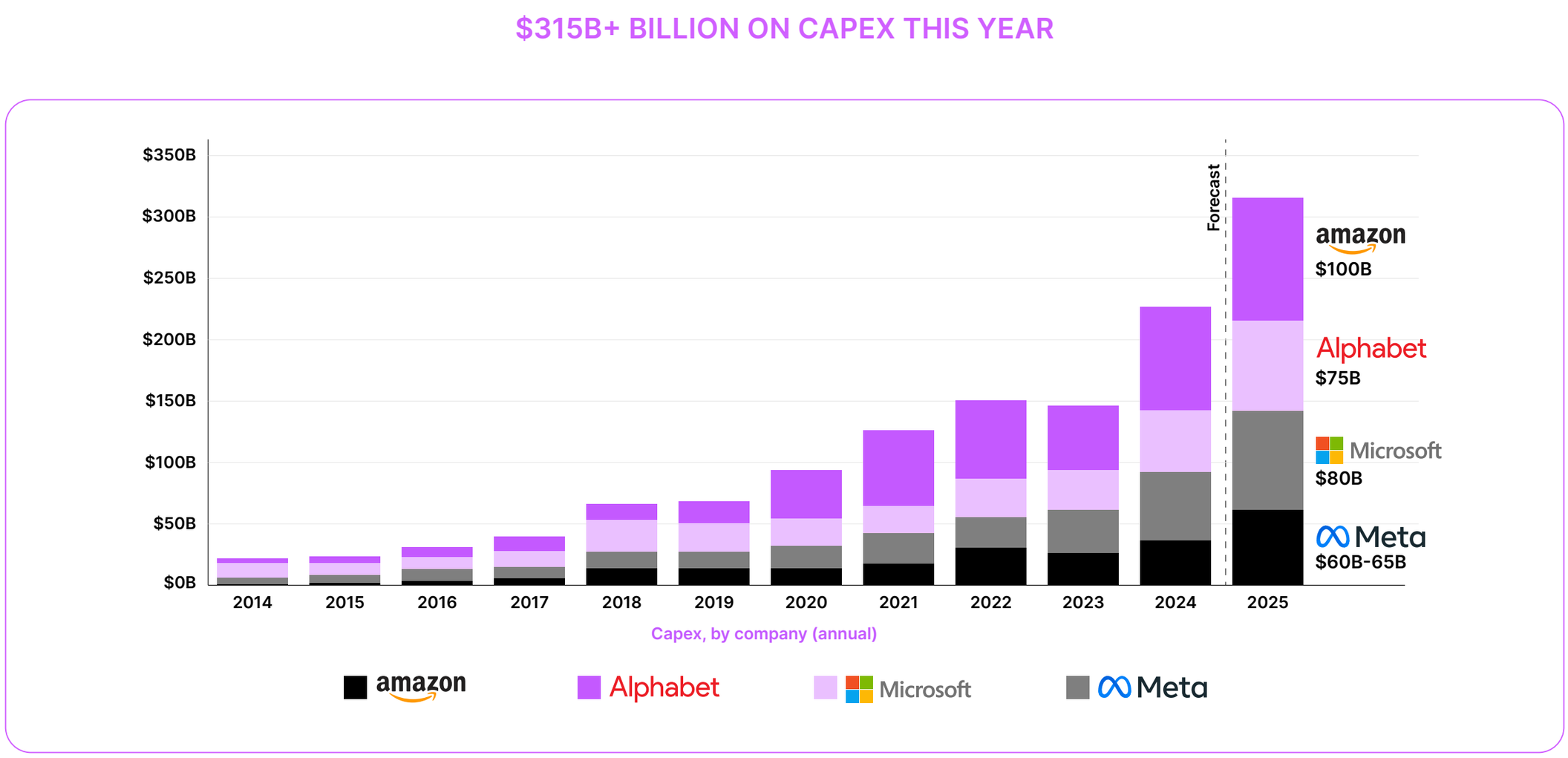
Source: Epoch.ai

The result is that aggregate spending on AI infrastructure continues to rise. Training clusters remain indispensable for each new generation, but inference fleets now account for the larger and faster-growing share of budgets, exerting increasing pressure on the AI supply chain.

The Capital Response and Market Bottlenecks

The scale of the response is most visible in capital expenditure. In 2025, Microsoft, Amazon, Google, and Meta together are guiding more than 300 billion dollars of capex, the majority directed toward AI-specific data centres.

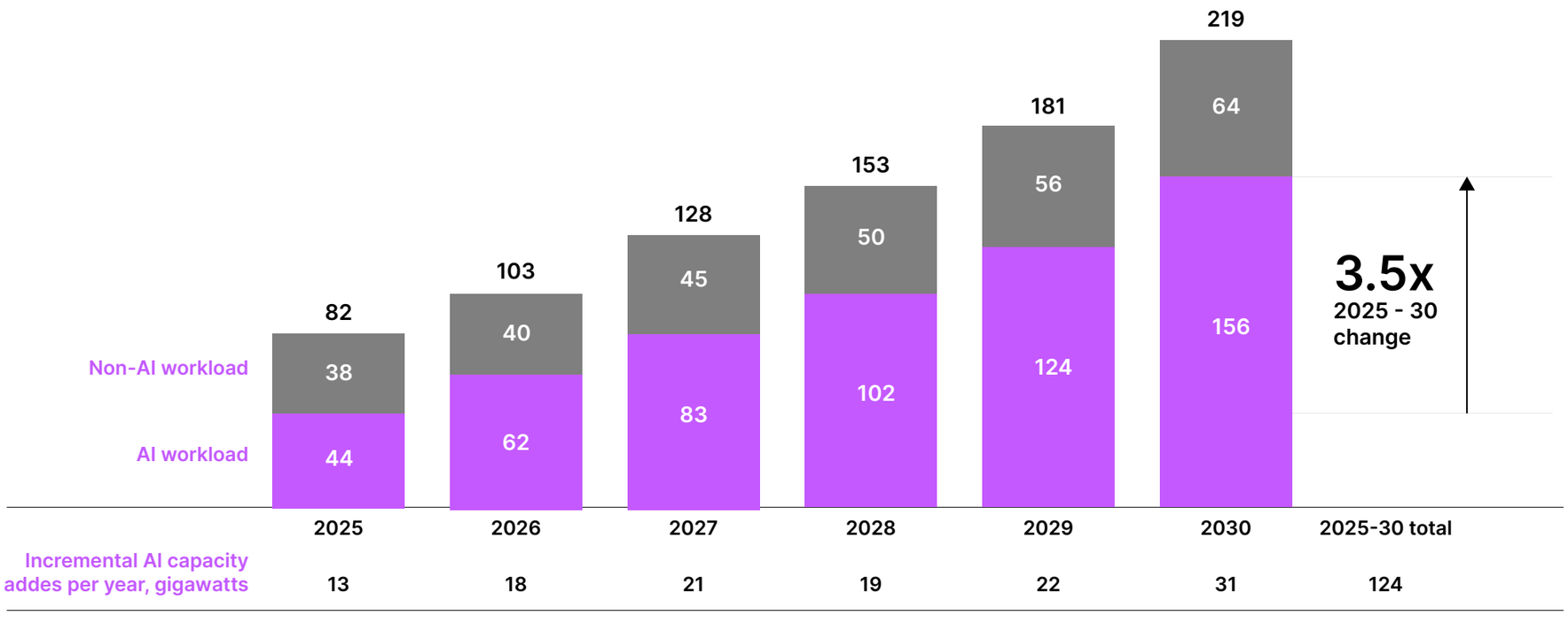
Microsoft has signalled roughly 80 billion, Alphabet 75 to 85 billion, Amazon more than 100 billion, and Meta 60 to 65 billion. These are not short-term IT budgets, but multi-year programmes designed to secure leadership in compute, networking, memory and power.



Source: Sherwood.news

Independent estimates underline how extensive this build-out will be. McKinsey projects that global data-centre investment will reach 6.7 trillion dollars by 2030, with 5.2 trillion linked directly to AI. Within this total, accelerators are the fastest-growing category.

BOTH AI AND NON-AI WORKLOADS WILL BE KEY DRIVERS OF GLOBAL DATA CENTER CAPACITY DEMAND GROWTH THROUGH 2030

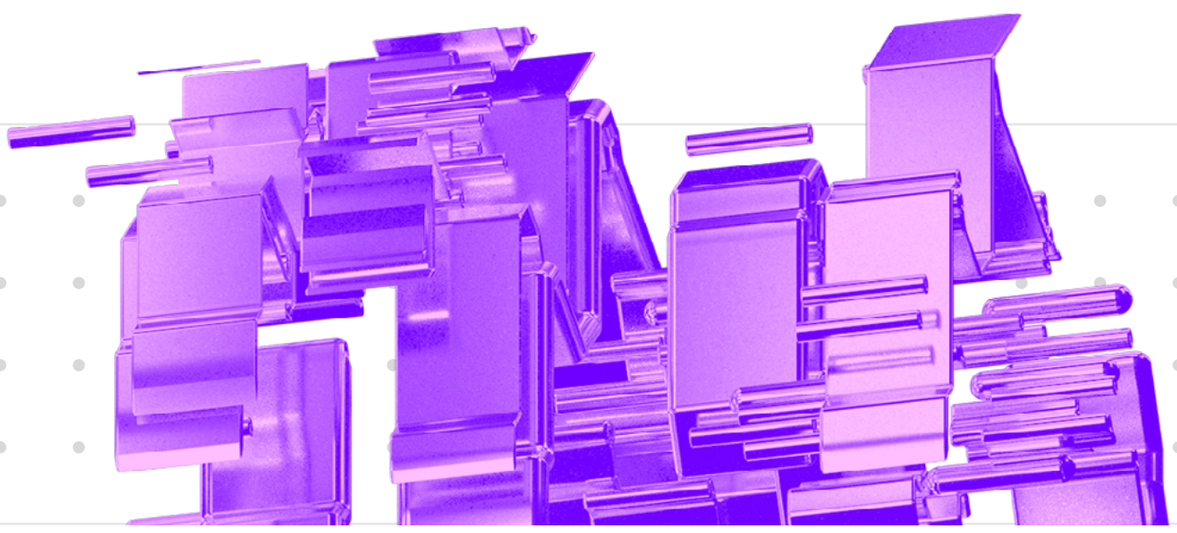


Source: Mckinsey

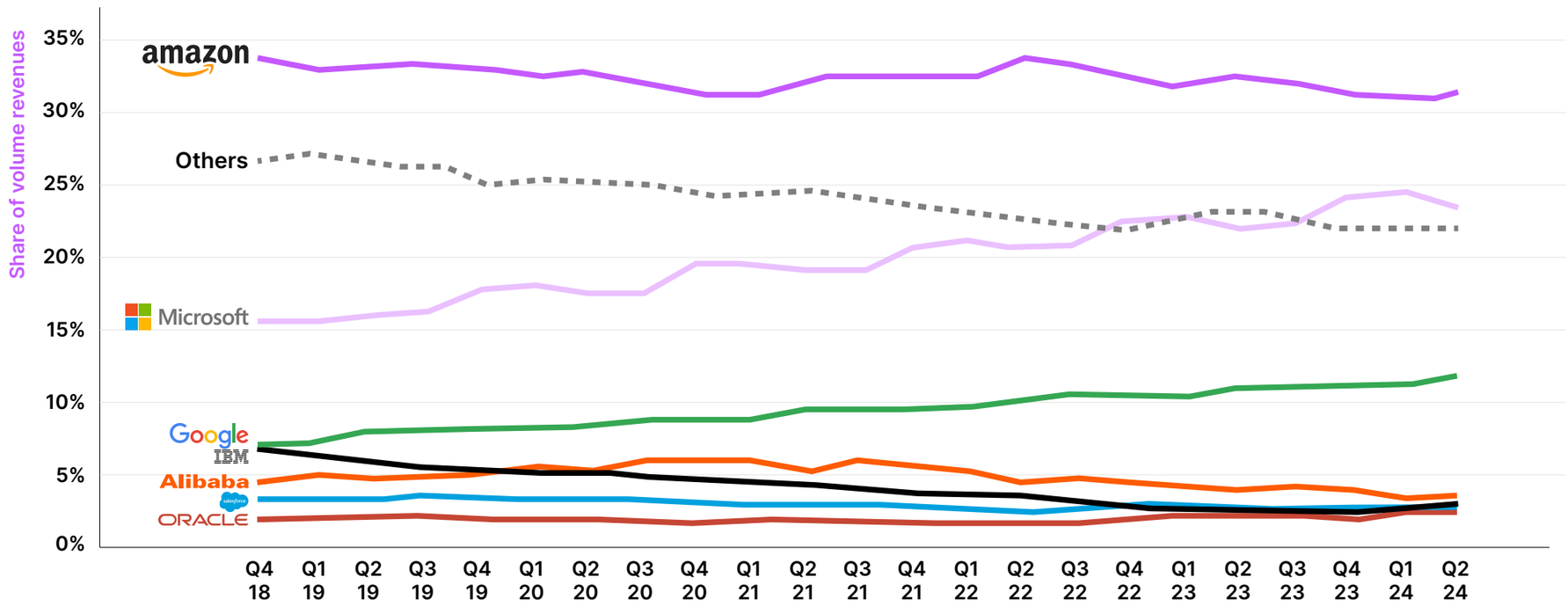
Storage and networking are also scaling rapidly. AI-related storage demand is projected to exceed 240 billion dollars per year by 2030, while high-performance networking for GPU clusters and retrieval systems is expected to surpass 100 billion.

These numbers also reveal the bottlenecks. At the silicon layer, supply is concentrated in TSMC's advanced nodes and packaging capacity, both of which remain constrained. Power is another critical limit. The International Energy Agency projects that global data-centre electricity use will more than double by 2030 to roughly 945 terawatt hours, with AI-optimised facilities driving most of the increase.

The structure of the market makes these constraints even harder to manage. A small group of hyperscalers such as AWS, Microsoft Azure and Google Cloud, together with Nvidia's DGX Cloud, controls most of the available capacity. Their pricing models, reservation systems and proprietary APIs make switching costly and keep customers locked into their ecosystems. Regulators are starting to respond.



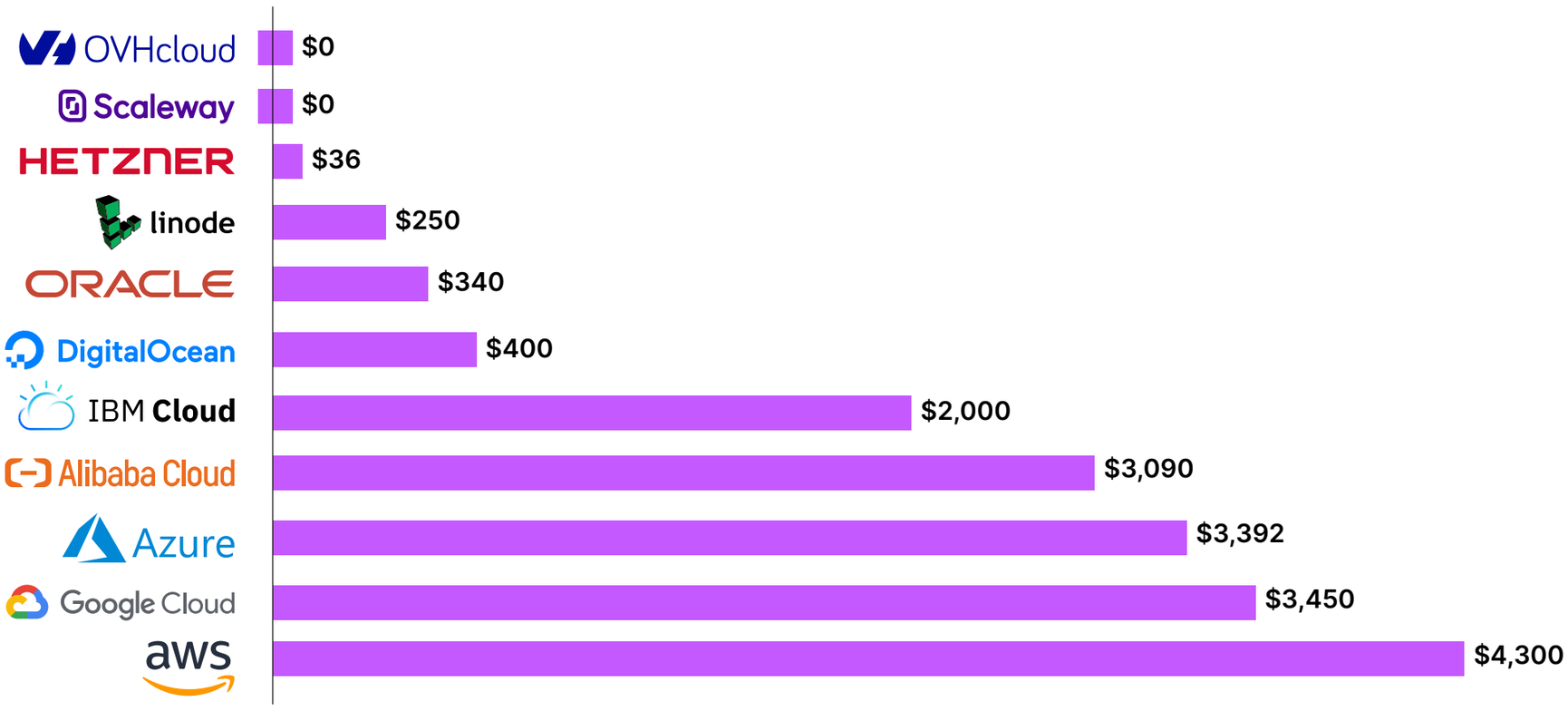
CLOUD PROVIDER MARKET SHARE TREND



Source: Holori

The UK Competition and Markets Authority has found that cloud infrastructure competition “is not working well,” citing egress fees, bundled services and long-term discounts as barriers to choice. The US Federal Trade Commission has raised similar concerns about preferential partnerships in the AI supply chain.

EGRESS COSTS FOR 50TB OF DATA



Source: Cast.ai

For startups, research labs and even governments, access to advanced compute increasingly requires long-term commitments to a handful of vendors.

The Case for Decentralising AI Infrastructure

Given the mounting bottlenecks in compute, storage and market access, there is a growing need to decentralise the foundations of AI.

Today's infrastructure is concentrated in the hands of a few hyperscalers, and this centralisation introduces risks that extend beyond commercial competition. It determines who controls access to AI, how data is handled, and which applications are prioritised.

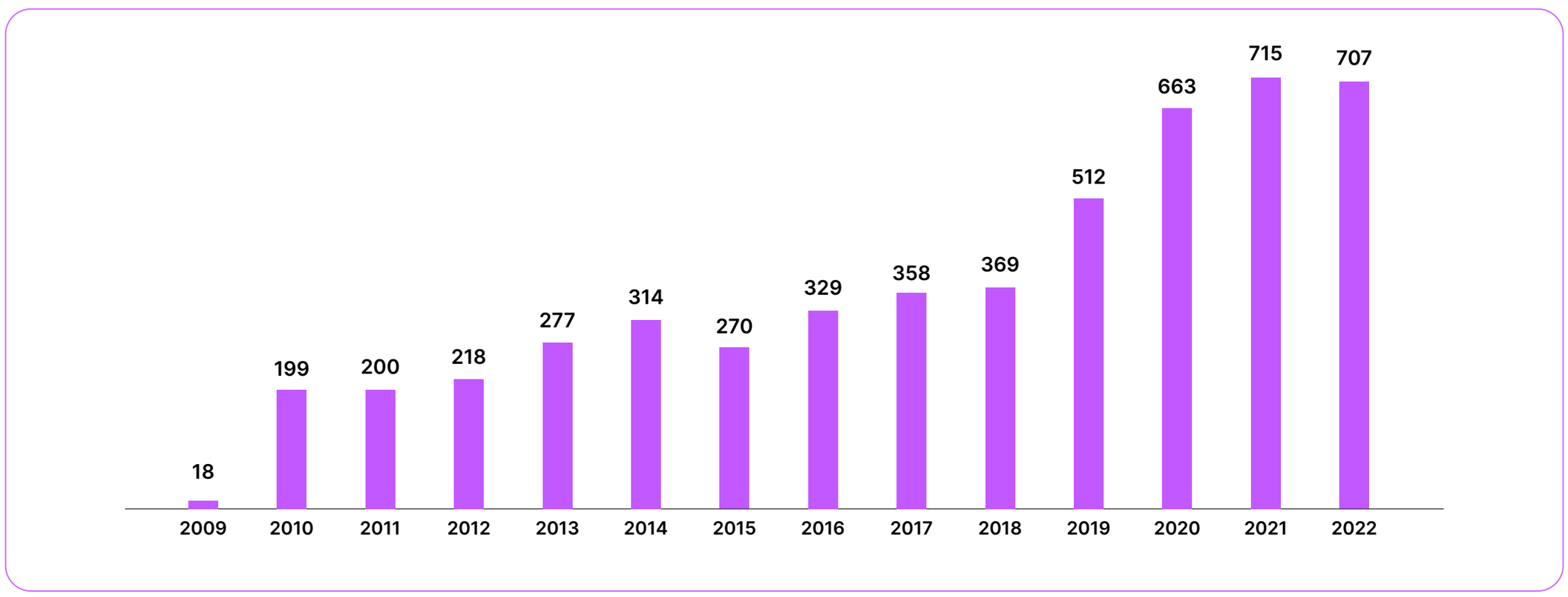
To build an AI economy that is resilient, accountable, and aligned with global interests, decentralisation must become a core design principle.

Privacy and Data Integrity

At present, nearly all training runs and inference requests are mediated by centralised platforms. Every interaction flows through systems owned and monitored by a handful of providers.

This creates persistent risks of misuse, from regulatory capture and commercial exploitation to outright data breaches. The stakes are highest in sensitive sectors such as healthcare, law and finance, where the data in question often contains personal identifiers or commercially critical information.

HEALTHCARE DATA BREACHES OF 500 OR MORE RECORDS



Source: Hipaa journal

Decentralised networks offer a different model. By opening compute and storage to permissionless marketplaces, they give users greater control over how data is processed and who can access it.

The benefits are not limited to cost reduction. They also provide stronger guarantees that personal or proprietary data will not be invisibly monetised or repurposed by intermediaries.

Transparency and Auditability

AI development today largely resembles a black box. Training datasets are opaque, provenance is rarely disclosed, and outputs cannot be independently verified. The problem is compounded when the same companies that supply GPUs also control the data pipelines and validation processes. Accountability in this model collapses into a trust-us framework.

Decentralised frameworks make verifiability a built-in feature. Provenance can be logged immutably, attribution can be tracked, and inference results can be challenged through mechanisms such as fraud proofs. In fields where correctness is non-negotiable, including science, finance and healthcare, this transparency is not a luxury but a requirement for adoption.

Alignment and Distribution of Power

The deeper challenge is structural alignment. When control of AI infrastructure rests with a handful of corporate boards and regulators, decisions about who gets priority access to GPUs or which applications are permissible are shaped by narrow commercial and geopolitical incentives.

Decentralisation offers an alternative alignment mechanism. By broadening participation in compute, storage and data availability, it redistributes power away from single points of control. Incentives can be designed to reward contribution and performance, not incumbency. The result is an AI infrastructure more closely aligned with a diverse set of global stakeholders rather than a select few.

Why Blockchain is a Fit

The constraints in AI are not only about hardware scarcity. They are coordination problems. Compute, storage, and bandwidth exist in large supply across data centres, enterprises, and individuals, yet these resources are fragmented and underutilised. Hyperscalers dominate because they centralise this capacity, bundle it into closed services, and enforce performance guarantees through proprietary contracts. Their control stems less from raw infrastructure and more from their ability to organise markets and enforce trust.

Blockchains provide an alternative coordination layer. They combine cryptographic verification with economic incentives to pool globally distributed resources into systems that are transparent, auditable, and resistant to capture. This makes them uniquely suited to address three structural needs of AI infrastructure: verifiability, resilience, and interoperability.

Verifiability

Correctness and provenance are essential for AI pipelines, yet today they depend entirely on trusting centralised providers. Blockchains make these guarantees enforceable.

- Data availability layers ensure that training sets, inference prompts, and results remain retrievable and auditable.
- Proof-of-access protocols require storage providers to show they genuinely hold the data they advertise.
- Zero-knowledge proofs make it possible to validate computations without revealing sensitive inputs.
- Fraud-proof mechanisms reward actors who detect incorrect results, creating adversarial incentives for integrity.

Together, these primitives convert AI infrastructure from a black box into a transparent system where claims can be independently verified.

Resilience

Centralised AI systems concentrate risk in a few data centres and supply chains. A decentralised model spreads resources across many participants, reducing single points of failure and making infrastructure more robust.

- **Distributed GPUs.** Idle or underutilised GPUs can be pooled into decentralised networks, creating a scalable marketplace for compute. This expands overall capacity and lowers barriers to entry compared to hyperscaler contracts.
- **Parallel workloads.** Sharding and partitioning allow large training and inference tasks to run across many nodes at once, improving throughput and efficiency.
- **Data durability.** Erasure coding disperses files into redundant fragments, ensuring availability even when individual providers fail.
- **Shared security.** Validator sets coordinate activity across networks, so failures in one area do not compromise the system as a whole.
- **Alignment through diversity.** By distributing compute across many providers, decision-making about access and usage reflects a broader set of stakeholders rather than a small group of corporations.

Together, these mechanisms support AI-scale workloads while keeping the system fault-tolerant, tamper-resistant, and more aligned with global interests.

Interoperability

Once compute and storage are verifiable, they can be linked natively with payments, identity, and governance. Blockchains act as the substrate for this integration.

- **Stablecoin rails** already clear more than $7 trillion annually, enabling global settlement of inference and training fees in real time.
- **Decentralised identity systems** such as World ID or Lit Protocol demonstrate how usage rights and licensing can be enforced transparently across networks.
- **Governance protocols** allow stakeholders to vote on resource allocation, ensuring infrastructure aligns with community and policy objectives rather than only corporate incentives.

By collapsing coordination, settlement, and governance into a shared state machine, blockchains make AI infrastructure programmable at a global level.

The Tech is Ready

These capabilities are no longer theoretical. Recent advances in blockchain design have made decentralised infrastructure viable for workloads at AI scale.

- Proof-of-stake consensus now achieves sub-second finality with thousands of transactions per second, proving that low-latency global coordination is possible.
- Zero-knowledge proof systems have seen costs fall by orders of magnitude, making the validation of complex computations efficient enough for practical deployment.
- Data availability has evolved from simple broadcast models to modular designs that use erasure coding, sampling, and parallel validation, enabling bandwidth at the level required for large training and inference datasets.

The convergence of these advances with rising demand for transparency and accountability creates a narrow but decisive window. Blockchains are not a peripheral add-on to AI. They are the technical foundation required to build infrastructure that is open, auditable, and resilient.

0G Ecosystem Overview

Now that we have established why blockchains are uniquely suited to coordinate AI infrastructure, the next step is to examine how 0G approaches this challenge. Most blockchain projects focus on a single part of the puzzle.

One chain may specialise in storage, another in compute, another in data availability. The result is a set of silos that developers must stitch together into fragile pipelines never designed to operate as one system. At AI scale, this fragmentation is unsustainable. Training and inference require storage, bandwidth, compute, and coordination to work seamlessly from the outset.

0G was built with this reality in mind. Rather than optimising for a single use case, it is designed from the ground up as a decentralised AI operating system (deAIOS).

Just as a conventional operating system abstracts hardware into usable services, deAIOS abstracts global pools of compute, storage, and bandwidth into verifiable and economically coordinated resources. The outcome is a modular yet integrated stack, where each layer is specialised but fully interoperable.

The deAIOS Vision

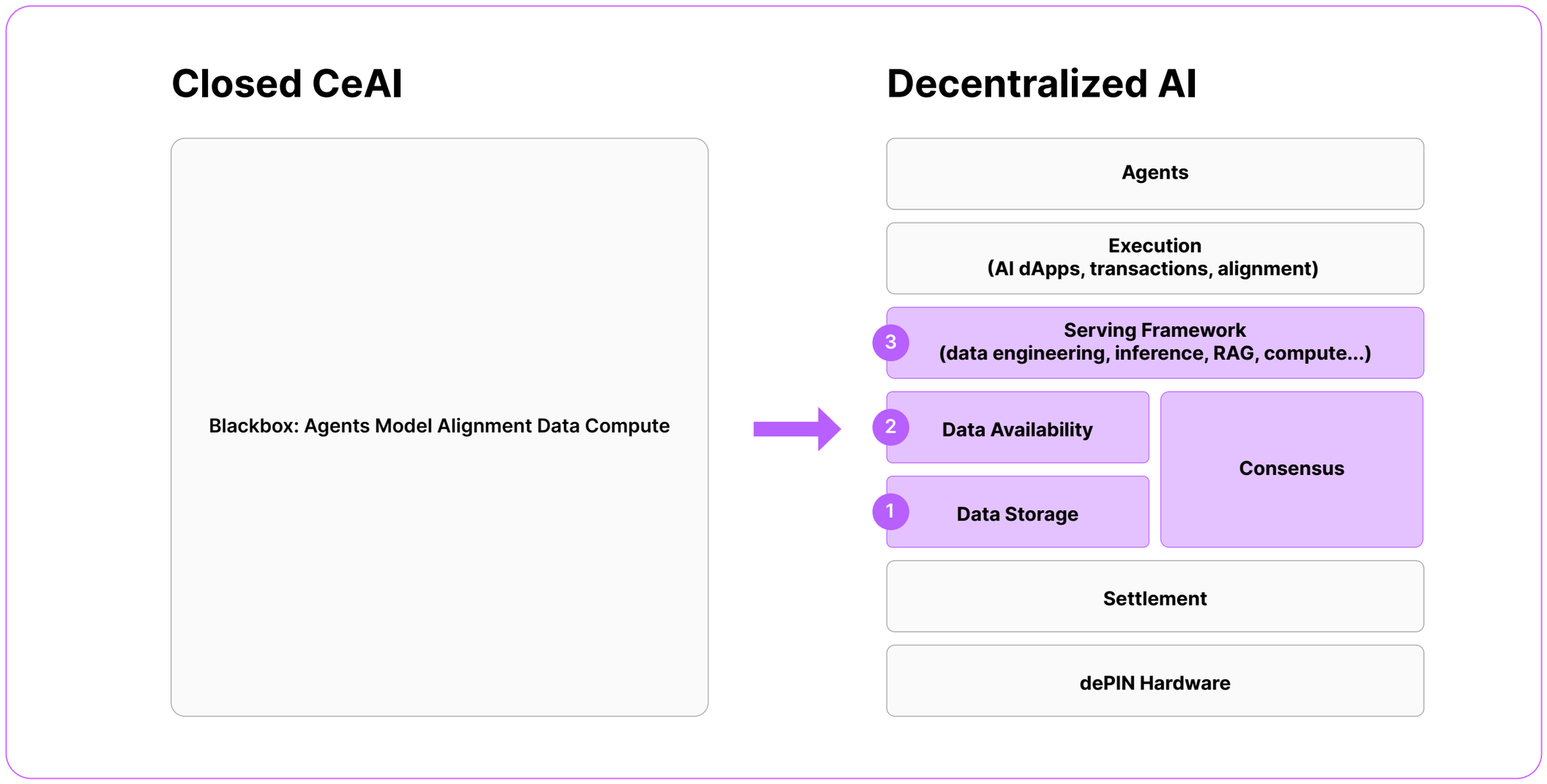
In the early days of computing, machines were programmed directly through a slow and complex process that only a handful of experts could manage. The development of the first operating system in the 1950s changed everything by automating repetitive tasks, managing memory, and providing a standard interface between hardware and users. This made computing more efficient and opened it to a much wider audience.

Until now, Web3 has lacked a comparable foundation for AI. Developers who wanted to build decentralised applications for training or inference had to stitch together storage protocols, compute providers, and bandwidth solutions. The result was fragile and fragmented.

0G introduces a different model. As the first operating system for decentralised AI, it coordinates core hardware resources in one integrated stack. By doing so it lays the groundwork for AI systems that are transparent, verifiable, and aligned with their users rather than locked inside corporate black boxes.

This design directly addresses the weaknesses of centralised AI. Data remains under user control rather than being harvested without consent. The provenance of training inputs and outputs can be verified rather than hidden. Contributors of data or compute are rewarded in open markets rather than excluded from value capture. And governance can be distributed among communities instead of concentrated in the hands of a few firms.

CLOSED CEAI vs DECENTRALIZED AI



Source: 0G

With this vision established, we can now look at how 0G structures its four main components.

Storage: Trust in Data

AI training depends on data provenance and persistence, but centralised storage makes these guarantees reputational rather than technical. 0G's storage layer enforces them cryptographically, combining immutable archival logs with flexible mutable workloads. Providers must prove they hold the data they claim, giving developers a reliable, verifiable training environment.

Data Availability: Removing Bottlenecks

AI workloads move more data than any general-purpose blockchain can handle. 0G solves this by separating payload lanes: large datasets are dispersed across partitions, while only lightweight commitments go through consensus. GPU-accelerated encoding and quorum sampling keep the system scalable, secure, and cost-efficient.

Compute: An Open Marketplace

Compute is the scarcest AI input, monopolised today by central providers. 0G reframes it as an open marketplace, where providers offer training or inference and settle transparently on-chain. This makes compute auditable, tradeable, and accessible globally — removing artificial scarcity and unlocking new forms of coordination.

Consensus: Scaling Security Horizontally

Tying it together is a multi-network consensus model. Multiple chains can run in parallel, each handling parts of the workload, while security is anchored in a root staking layer. This horizontal model scales throughput without fragmenting security, and can interoperate with external restaking frameworks to borrow additional guarantees.

Why 0G Succeeds Where General-Purpose Blockchains Fail

The gap between 0G and a monolithic chain like Solana can be illustrated with a simple analogy. Solana is like a single house: everything happens under one roof, but there is only so much space before it becomes overcrowded. 0G, in contrast, is more like an apartment complex. Each unit is designed for a specific function, yet all share the same foundation. The system can grow by adding more units without straining the structure as a whole.

0G vs SOLANA

This difference is crucial. General-purpose blockchains were never built to manage the bandwidth of AI-scale storage, the partitioning needed for true availability, or the verification systems that make computation auditable. 0G tackles these bottlenecks directly. Its modular and purpose-built design gives AI the infrastructure it has been missing: scalable data flows, trustworthy storage, accessible compute, and unified security.

In short, 0G is not just another blockchain optimised for one task. It is a complete operating system for decentralised AI, designed to support workloads at trillion-dollar scale and to make verifiable, transparent, and user-aligned AI a reality.

Technical Deep Dive

Up to this point, we have explored 0G from a high-level perspective, focusing on its mission and vision as a decentralised operating system for AI. To truly understand how this vision is realised, however, we need to examine the underlying technology.

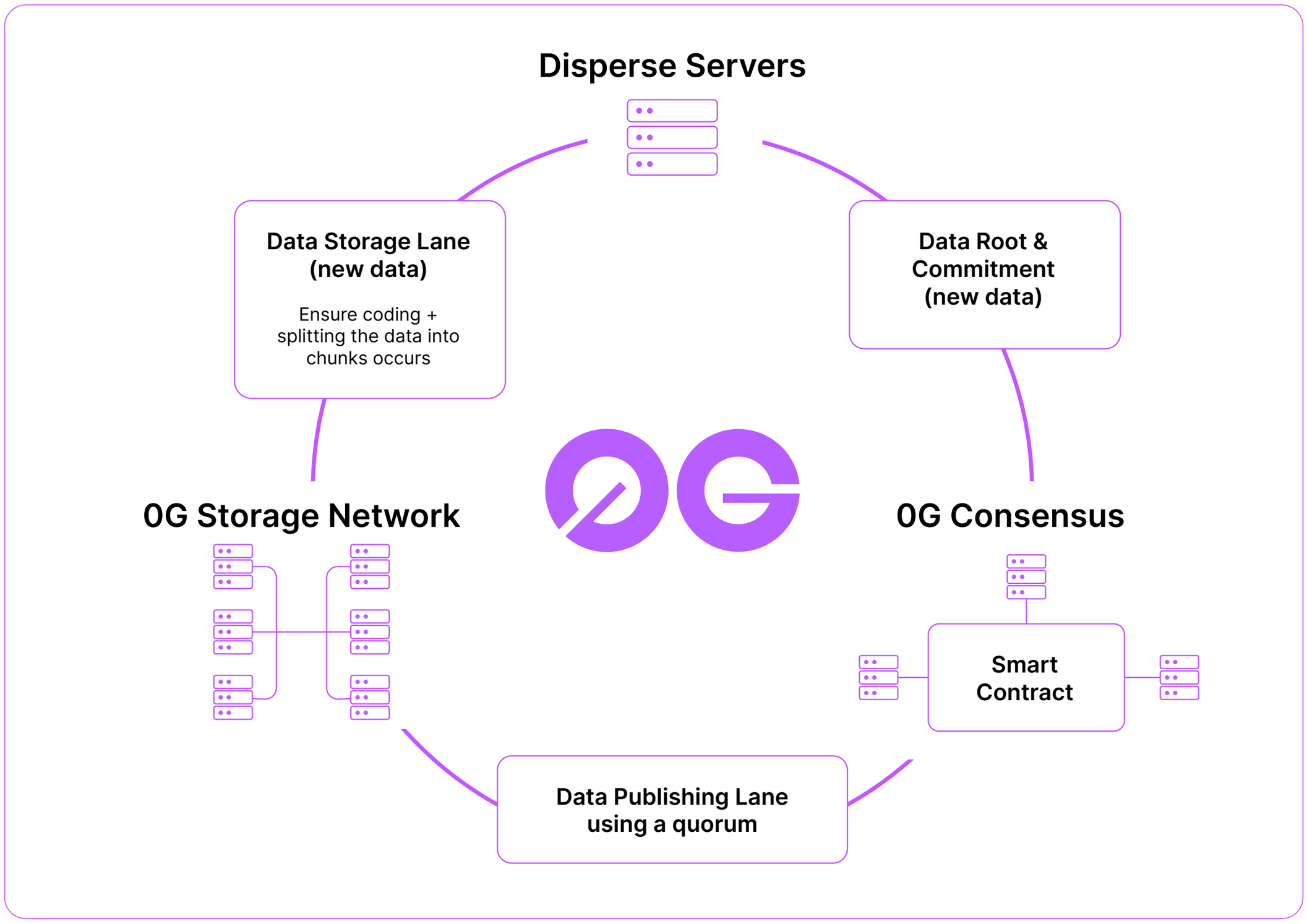
Each component of 0G's stack, from storage and data availability to compute and consensus, has been designed with a specific role in enabling scalable, verifiable, and economically sustainable AI. By breaking down these layers, we can see how the system functions in practice and how its design choices address the limitations of general-purpose blockchains.

0G Storage

Storage is the foundation of any decentralised AI infrastructure because AI workloads are uniquely data-hungry. Training large-scale models requires terabytes of raw corpora that must be stored immutably, while inference and application-level tasks demand low-latency access to dynamic and constantly changing data. A system that can only handle one of these use cases will inevitably break down at scale.

0G Storage addresses this challenge through a two-layer architecture that separates permanent from mutable data. The Log Layer functions as an immutable, append-only archive. Once data is written, it cannot be altered, making it well suited for large files such as training datasets, high-volume telemetry, image or video repositories, or blockchain history. For AI developers, immutability provides a critical guarantee of provenance: the ability to prove which dataset was used to train a model. This strengthens reproducibility and compliance while ensuring accountability when model behaviour needs to be traced back to its source data.

The Key–Value Layer, in contrast, is designed for workloads that evolve in real time. Optimised for fast updates and key-based retrieval, it can support agent memory that grows with interaction, databases that demand constant changes, or backends like game states and collaborative documents. By combining both layers in one system, 0G avoids the need for developers to juggle separate storage solutions for archival and dynamic data.

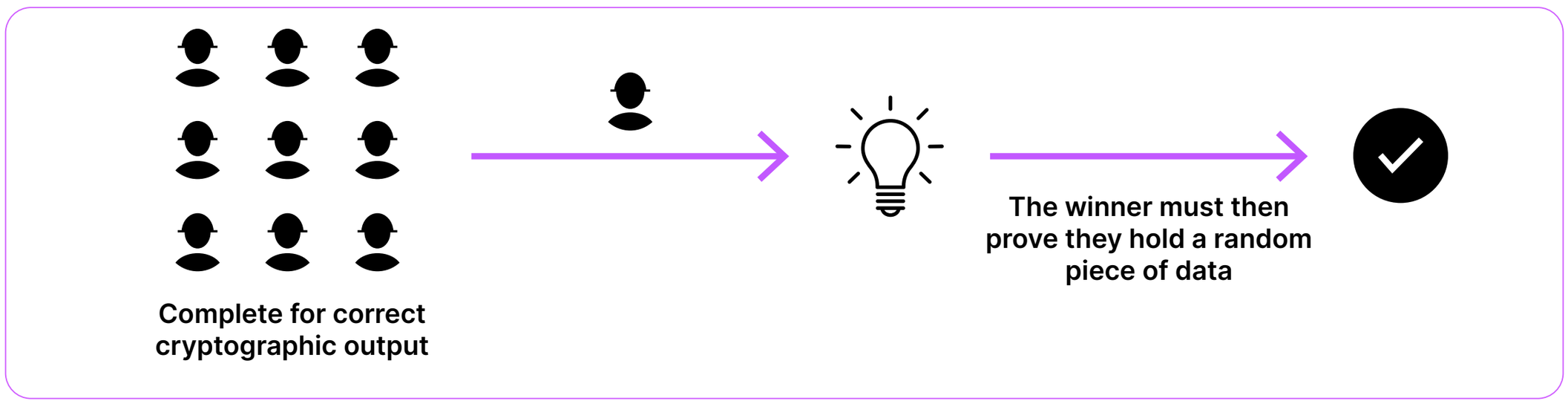


Source: 0G

Security and Incentives

The integrity of these layers is guaranteed by Proof of Random Access (PoRA), a consensus mechanism that forces providers to prove they are storing what they claim. At random intervals, miners are challenged to fetch a 256 KB chunk from their committed dataset.

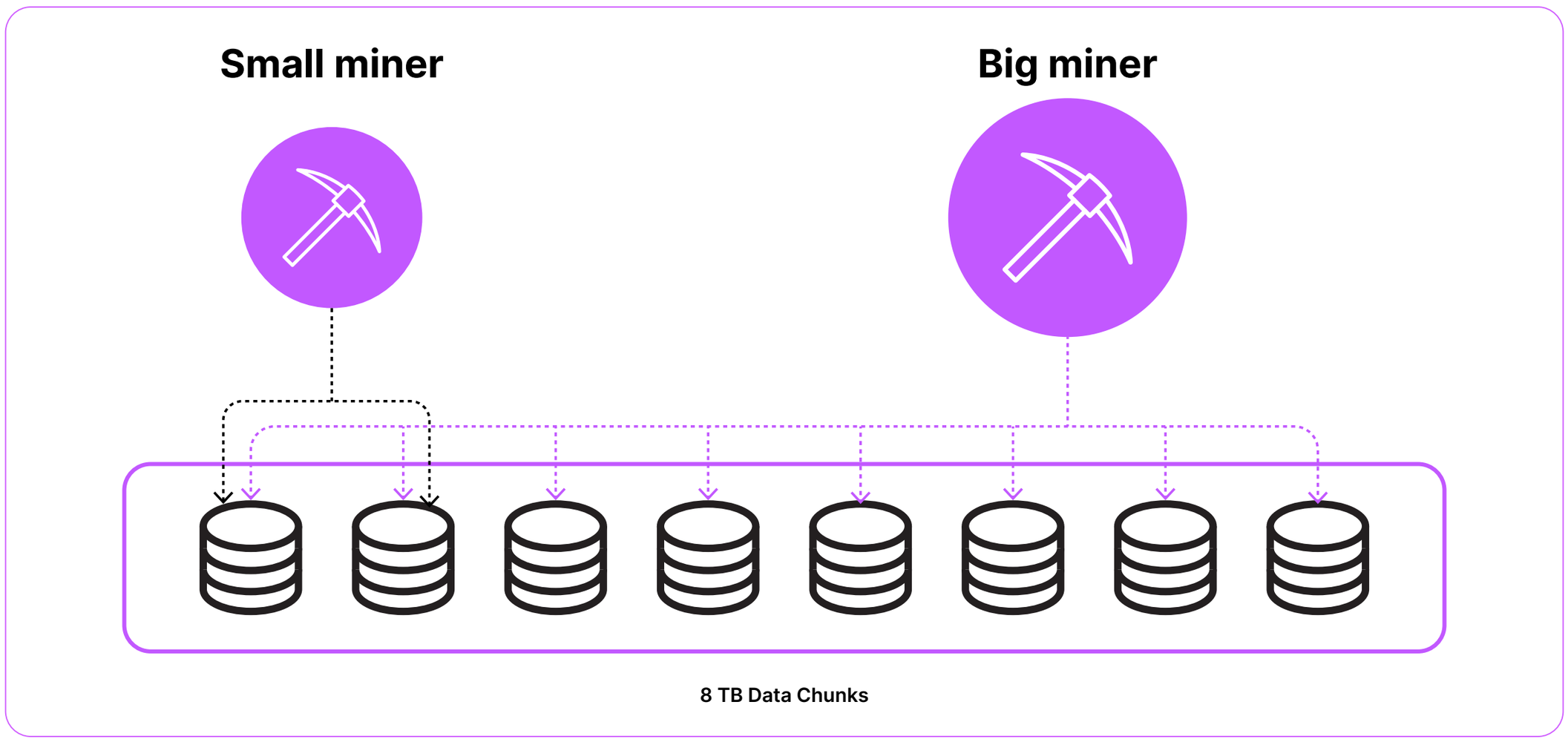
0G'S PROOF OF RANDOM ACCESS



Source: 0G

The retrieval must be performed quickly and verifiably; those who succeed earn rewards, while those who fail are excluded. To prevent industrial-scale dominance, each provider's commitment is capped at 8 TB per "mining range." This cap levels the playing field by ensuring that smaller providers can compete effectively, making the network more decentralised and resistant to capture.

SMALLER MINERS CAN COMPETE WITH BIGGER MINERS



Source: 0G

Unlike mining systems that reward raw processing power, PoRA is I/O-bound rather than compute-bound. Success depends on being able to retrieve data rapidly, not on the size of a provider's hardware farm. This aligns incentives directly with user needs: what matters is availability and responsiveness, not brute-force scale.

Economically, 0G uses an endowment model. Users pay upfront for the duration they want their data stored, and providers earn rewards gradually as long as they continue to prove storage. This creates predictable pricing and avoids the "race to the bottom" dynamics that undermine many decentralised networks. It also embeds self-balancing incentives: if a file is under-replicated, its reward rate rises, encouraging more providers to pick it up until redundancy is restored.

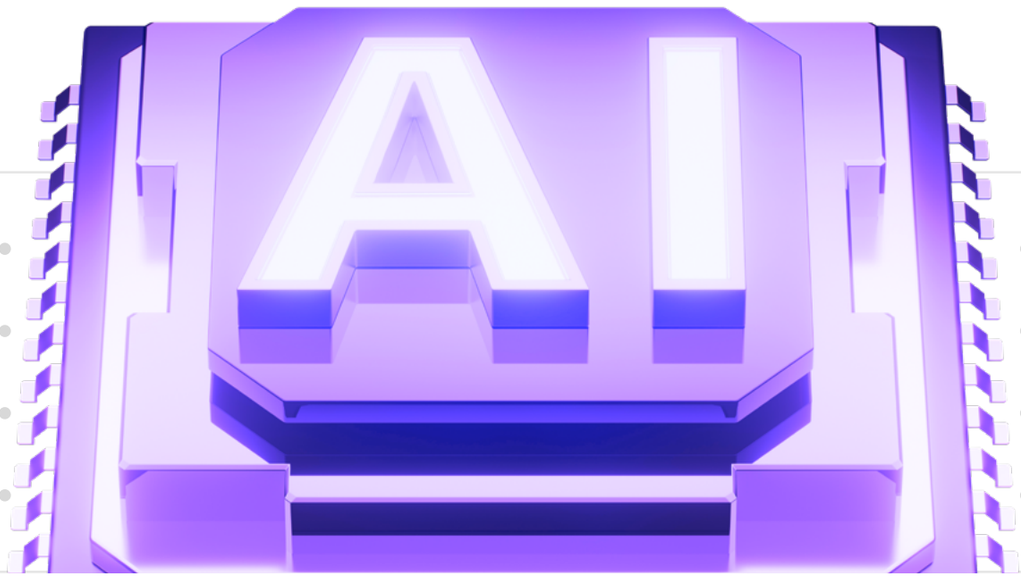
Taken together, these features make 0G Storage more than just a persistence layer. It combines immutability for long-term datasets, flexibility for real-time applications, and verifiability at the protocol level. For developers, this creates a backbone capable of handling AI-scale workloads with confidence. For providers, it establishes an economic model where reliability and fairness are rewarded, rather than sheer hardware dominance.

ERC-7857: Privacy-Preserving Ownership

A core challenge in digital asset infrastructure is defining ownership for intelligent or high-value models in a way that is both secure and verifiable. Existing NFT standards such as ERC-721 and ERC-1155 store only identifiers and metadata references, leaving the actual models, data, and behavioural logic off-chain.

This separation creates a gap between symbolic and functional ownership. When a token changes hands, the underlying intelligence often does not, resulting in incomplete or non-verifiable control.

To address this, 0G introduced ERC-7857, a protocol designed to unify ownership, privacy, and verifiable transfer. It extends the NFT model with encrypted metadata and re-encryption mechanisms that ensure sensitive information travels securely with the asset itself. Each token contains an encrypted representation of the underlying intelligence or dataset, enabling functional ownership without revealing proprietary content.



KEY FEATURES OF ERC-7857

FEATURE	ERC-721	ERC-7857
Metadata Type	State and public	Dynamic and private
iNFTs	Not supported	Supported
Storage	Typically, a URI pointing to a JSON file on IPFS or centralized servers	Encrypted metadata stored securely on decentralized platforms like 0G Storage
Metadata Transfer	Only the transfer token ID is; metadata access is not included	Ownership and metadata (e.g., neutral models, memory) are securely transferred together
Privacy and Encryption	No native support for encryption or private storage	Full support for encrypted metadata and privacy-preserving transfers
AI-Specific Applications	Limited applicability for AI agents	Tailored for AI agents, supporting lifecycle management, ownership verification, and more
Integration with Decentralized Storage	Limited or requires manual integration	Seamless integration with decentralized storage like 0G Storage
Key Use Cases	Digital art, collectibles, gaming assets	AI marketplaces, AI-as-a-Service (AIssS), enterprise AI ownership, IP monetization

Source: 0G

To achieve this, ERC-7857 introduces a set of capabilities that redefine how digital assets are secured, transferred, and updated. These features ensure that ownership remains both verifiable and private, even as assets evolve over time:

- **Encrypted metadata** that protects model weights and configurations within the token.
- **Secure re-encryption** that allows asset transfers to occur without exposing any underlying data.
- **Verifiable transfer** supported by Trusted Execution Environments (TEE) or Zero-Knowledge Proofs (ZKP) to confirm authenticity and integrity.
- **Dynamic metadata** that enables assets to evolve while maintaining a continuous and verifiable cryptographic identity.

Together, these features establish a more complete and trustworthy ownership model than existing standards. Frameworks such as x402 and ERC-8004, as well as other known systems like Virtual ACP, Google A2A, and Stripe ACP, each address specific aspects of digital asset management. Some focus on payments, others on identity or intellectual property protection. However, all of them leave the underlying data layer exposed or dependent on external systems.

ERC-7857 closes this gap by introducing native privacy and verifiability directly on-chain. It is the only framework that combines encryption, secure re-encryption, and dynamic metadata within a single protocol, ensuring that both ownership and the underlying intelligence remain private, portable, and cryptographically proven.

ERC-7857 vs OTHER STANDARDS

FEATURE / STANDARD	MCP	x402	STRIPE ACP	GOOGLE A2A	ERC-8004	VIRTUAL ACP	0G EIP-7857
Launch Date	November 2024	May 2025	September 2025	April 2025	August 2025	February 2025	Live now
Programming Language	Python, TypeScript, C#, Java	TypeScript, Python	JavaScript, Ruby, Python	Python, TypeScript	Solidity	Python, TypeScript	Solidity
Open Source	✓	✓	✓	✓	✓	✓	✓
Blockchain Based	✗	✓	✗	✗	✓	✓	✓
Payment Focused	✗	✓	✓	✗	✗	✓	✗
Commerce Focused	✗	✗	✓	✗	✗	✓	✓ 0G Storage + DA
Data Integration	✓	✗	✗	✗	✗	✗	✓
Agent Communication	✗	✗	✗	✓	✓	✓	✓
Supports Tool Integration?	✓	✗	✓	✗	✓	✓	✓
Privacy/ Encryption Features	✗	✗	✗	✓	✓	✗	✓ UNIQUE: TEE/ZKP encryption
Latency	Low	Low	Low	Low	High	High	Very Low
Underlying Mechanism	HTTP-based, ms range; reduces latency by up to 85% for long prompts	HTTP-based, ~200ms end-to-end, instant	HTTP-based, ms range; depends on commerce backend	HTTP-based, up to 40% latency reduction in agent comms	Blockchain-based, seconds to minutes due to on-chain confirmations	Blockchain-based, seconds to minutes due to on-chain confirmations	Blockchain-based, sub-second finality 0G: 50 Bgps DA+ sub-second consensus
GitHub Stars	6K	2.3K	775	20.3K	N/A	17	N/A

Source: 0G

When combined with 0G's high-throughput consensus and decentralized storage, it achieves sub-second finality and large-scale encrypted data transfer without compromising confidentiality. This architecture makes ERC-7857 practical for real-world applications that demand both security and scalability.

In practice, ERC-7857 can support a wide range of use cases across digital and AI-driven ecosystems.

- **Decentralized finance:** Trading agents and algorithmic strategies can be tokenized and transferred securely, allowing market participants to exchange functional models without exposing proprietary code or data.
- **Personal AI systems:** User-trained assistants can be reassigned, licensed, or shared with trusted parties while keeping all personal or behavioral information encrypted.
- **Gaming and simulation:** Non-player agents and digital entities can be traded as evolving assets that retain their behavioral history and skill progression under verifiable ownership.
- **Creative and industrial applications:** Artists, researchers, and enterprises can distribute trained models or generative tools as transferable assets, maintaining full control over how and where they are used.

ERC-7857 therefore extends beyond simple digital collectibles to form a foundation for a new category of intelligent, privacy-preserving assets. Its design aligns secure data handling with on-chain verifiability, providing the infrastructure required for scalable and trustworthy AI-driven economies.

From a research and infrastructure perspective, ERC-7857 is one of the first practical frameworks to bring together encryption, verifiability, and asset evolution within a single standard.

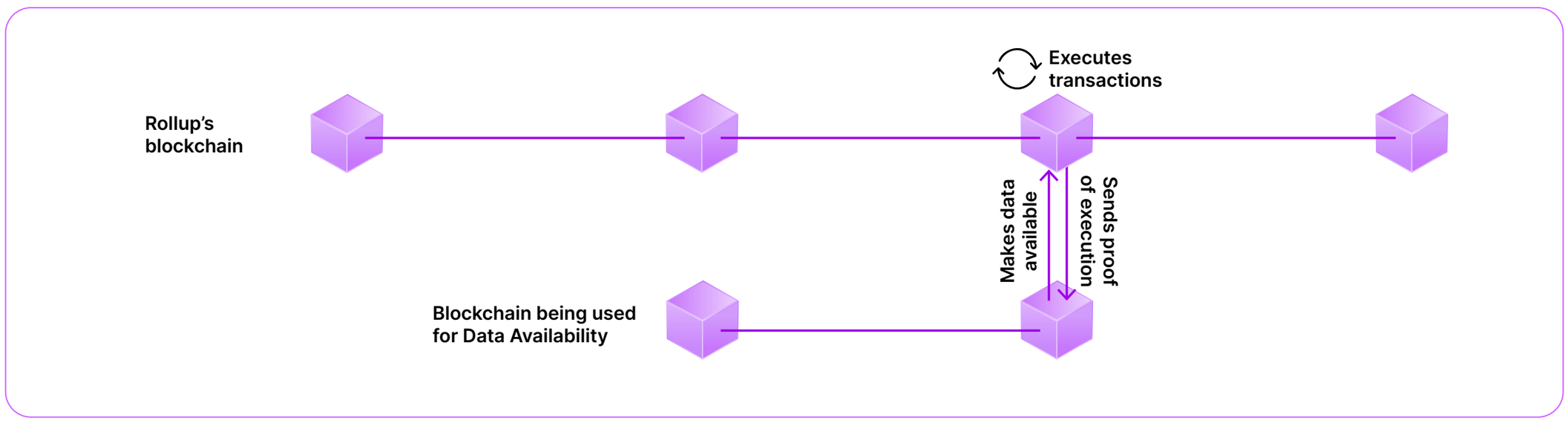
Its integration with the 0G ecosystem, including 0G Storage, Data Availability, and Compute, shows how cryptographic privacy can operate alongside high-performance infrastructure. This combination positions 0G's architecture as a credible foundation for decentralized AI ownership and exchange.

Data Availability (DA)

Data availability (DA) is one of the least visible but most critical components of decentralised infrastructure. It refers to the guarantee that data submitted to a network is not only published, but can also be accessed, verified, and retrieved by anyone who wishes to audit it. Without DA, decentralisation breaks down: validators cannot independently verify the state of the system, censorship cannot be detected, and fraud proofs cannot function.

In practice, DA is what makes scaling possible. Rollups such as Arbitrum or Base publish their transaction data back to Ethereum so that anyone can verify the rollup's history, but doing this directly on Ethereum is costly and bandwidth-constrained. DA layers emerged to offload this burden, offering lower-cost, higher-throughput infrastructure that ensures the same verifiability.

ROLLUP USING ANOTHER BLOCKCHAIN FOR DA



Source: Avail

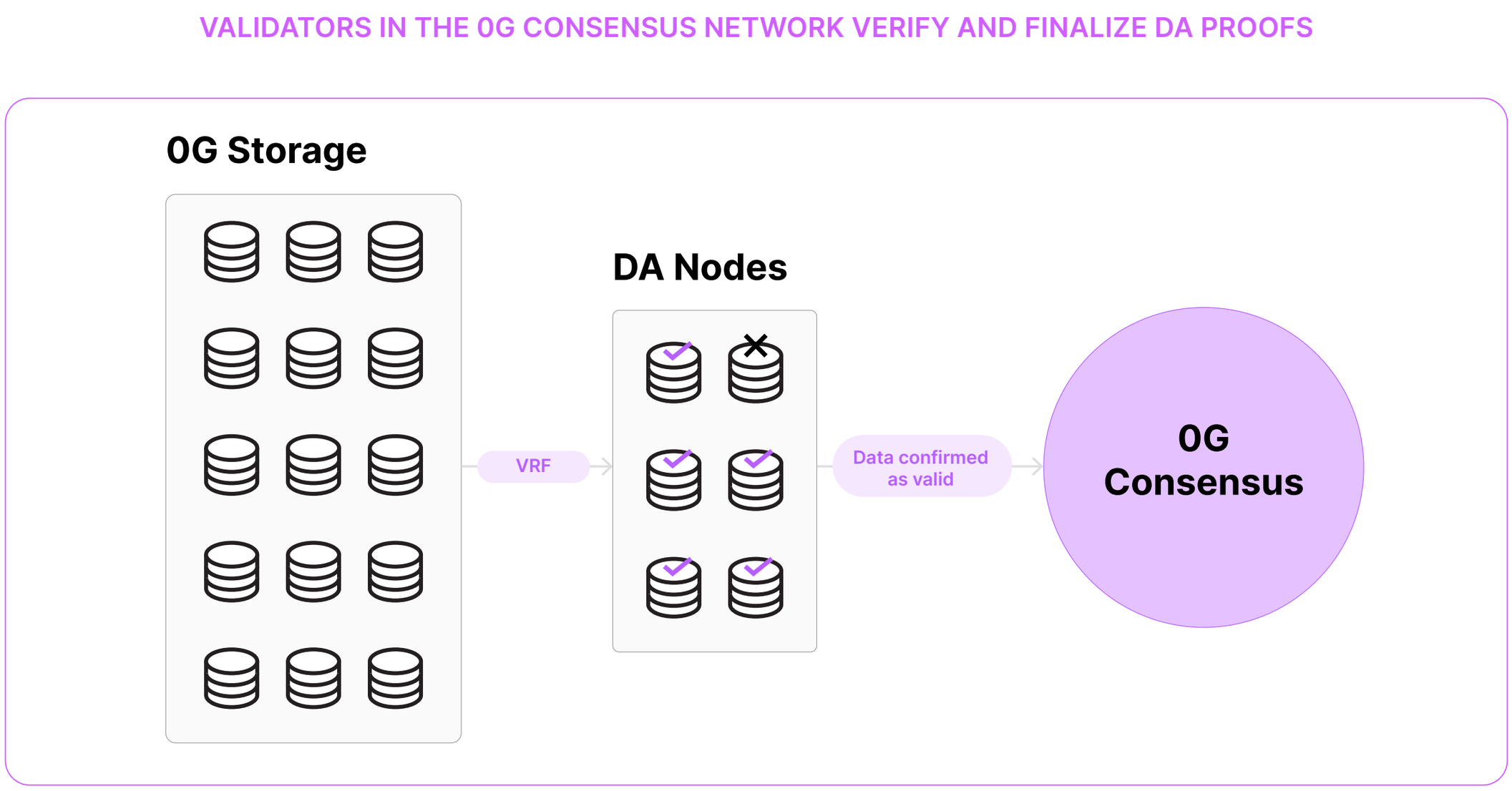
The problem is that most DA layers still rely on broadcasting large amounts of data to all participating nodes, which limits scalability to the slowest node in the network. They also lack integrated storage, meaning they depend on external systems for persistence. This creates inefficiencies in cost, retrieval speed, and reliability.

0G takes a different approach. Its DA module is designed from the ground up for AI-scale workloads: massive datasets, frequent queries, and strict performance requirements. Rather than trying to stretch general-purpose DA models to fit AI, 0G treats DA as a first-class citizen within its broader operating system.

0G's Design and Differentiation

The foundation of 0G's DA is its integration with 0G Storage. Data is erasure-coded and split into chunks that are distributed across a decentralised network of storage nodes.

To confirm availability, specialised DA nodes are randomly assigned using a Verifiable Random Function (VRF). This prevents collusion by making group membership unpredictable and auditable. These nodes then work in quorums to sample stored data and verify its presence, drastically reducing the amount of data that needs to be checked while still preserving strong security guarantees.



Source: 0G

When a quorum confirms that data is available, it submits an availability proof to the 0G consensus network. Validators, who are distinct from DA nodes, verify and finalise these proofs.

Security is anchored in a shared staking model: validators stake tokens on a root network (likely Ethereum), and slashable events across any connected subnet are enforced at this root layer. This ensures that security scales horizontally without fragmenting capital, while also allowing 0G to inherit Ethereum's economic security base.

Performance is where 0G's DA system distinguishes itself. Instead of requiring every node to process every piece of data, its sampling-based approach allows the network to scale indefinitely.

New consensus networks can be added as demand increases, and throughput grows horizontally without overloading individual nodes. Benchmarks on the Galileo Testnet already demonstrate throughput in excess of 50 Gbps, showing that the system can handle not just blockchain transactions, but the massive data flows demanded by AI training and inference.

The economic incentives reinforce resilience. As with storage, DA nodes are rewarded for correct and timely participation, while under-replicated or under-served datasets automatically attract higher rewards until redundancy is restored. This makes the system self-correcting and resistant to degradation over time.

The result is a DA layer that combines efficiency, scalability, and verifiability. For rollups and blockchains, it provides a high-throughput substrate that reduces costs and accelerates settlement.

For AI developers, it creates an environment where large datasets can be made accessible and auditable at scale, enabling everything from training language models to coordinating agent networks. By embedding DA directly into its operating system, 0G avoids the inefficiencies of bolted-on solutions and establishes itself as a backbone for decentralised intelligence.

Compute Network

If storage is the foundation of AI infrastructure, compute is its most expensive and scarce resource. Training or even fine-tuning advanced AI models requires specialised hardware, primarily GPUs. Access to these resources is dominated by a handful of centralised providers such as AWS, Google Cloud, or Azure, which impose high fixed costs, restrictive contracts, and vendor lock-in.

Smaller developers and startups face steep barriers: monthly GPU costs often run into the tens of thousands, API services charge per request at rates that quickly add up, and building in-house infrastructure requires millions in upfront investment. The result is that meaningful access to AI computing remains concentrated in the hands of well-funded institutions.

0G Compute seeks to break this bottleneck by reframing compute as an open marketplace. Instead of renting capacity from a few centralised providers, developers can access a global pool of GPU owners who contribute idle or dedicated hardware.

FEATURE	TRADITIONAL CLOUD	0G COMPUTE
Pricing Model	Fixed monthly costs	Pay-per-use
Provider Options	Limited vendors	Global GPU network

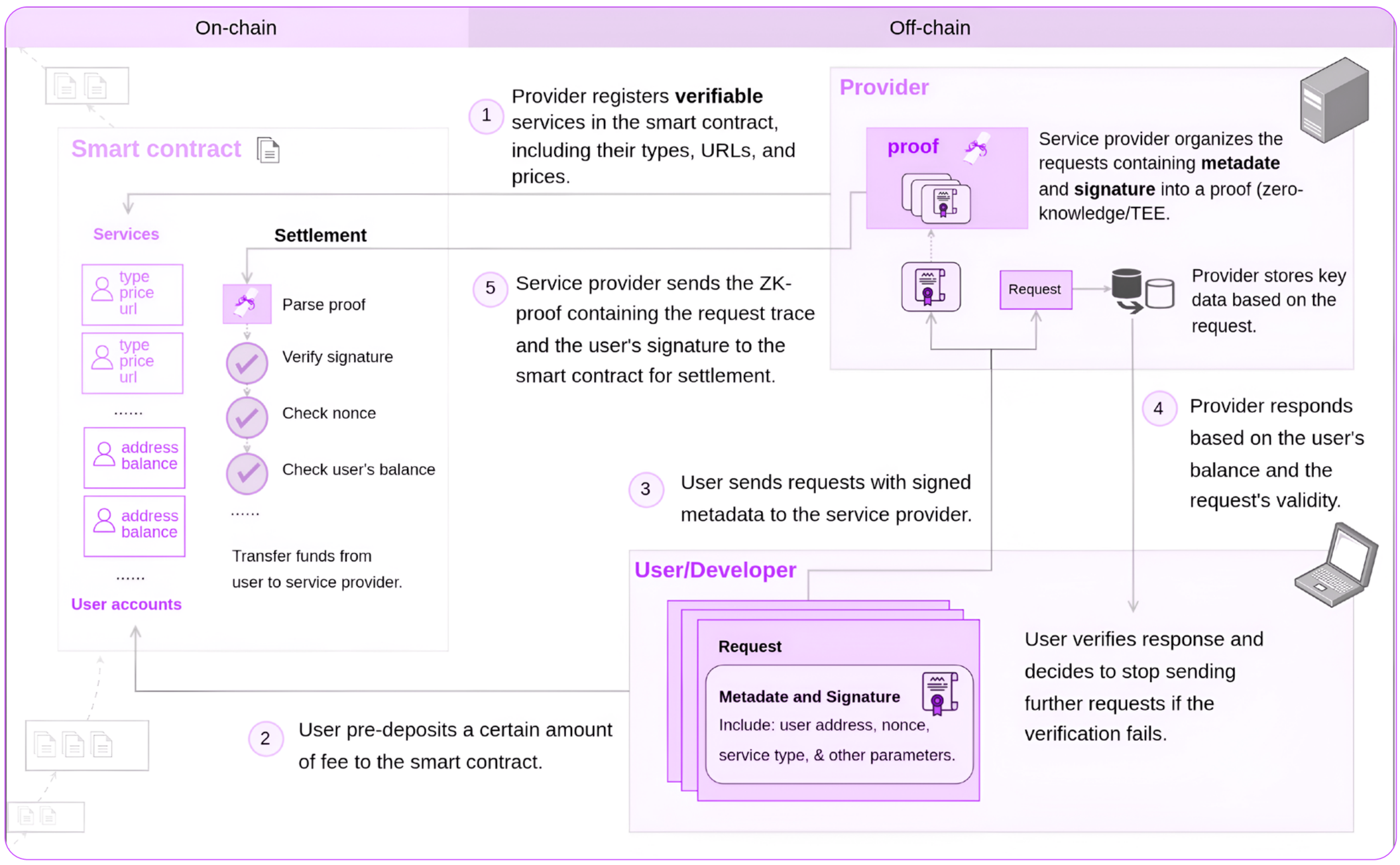
In this model, supply and demand are matched dynamically, and developers pay only for the compute they actually use. The promise is both cost efficiency, often quoted as up to 90 percent cheaper than cloud incumbents, and a more open system that lowers the barriers to participation in AI development.

Architecture and Incentives

The 0G Compute Network is structured around three key roles: users, providers, and the verification layer. Developers or AI users pre-fund their accounts and submit requests for inference, fine-tuning, or eventually full model training. These requests are automatically routed to the most suitable available GPU provider. Once a task is completed, payment is released through a smart contract escrow, ensuring trustless settlement.

For GPU owners, the process is equally straightforward. Providers register their hardware, set availability and pricing, and receive jobs through the network. Successful completion of tasks results in immediate earnings, creating a new revenue stream for otherwise idle or underutilised GPUs. This market design has the potential to aggregate resources at scale, ranging from enterprise-grade clusters to individual consumer GPUs, without requiring central coordination.

ARCHITECTURE OVERVIEW



Source: 0G

Trust and verification are central to the design. Computation is accompanied by cryptographic proofs that confirm the work was carried out correctly. Techniques such as trusted execution environments (TEE) and zero-knowledge proofs can be used to guarantee that outputs are accurate and tamper-proof. This prevents providers from faking results and ensures that users can rely on the integrity of the computation.

The incentive system aligns naturally with this architecture. Payments flow only after proof of work is verified, eliminating the risk of providers being paid for incomplete or incorrect jobs. At the same time, the global nature of the marketplace drives competition, pushing down costs while broadening availability.

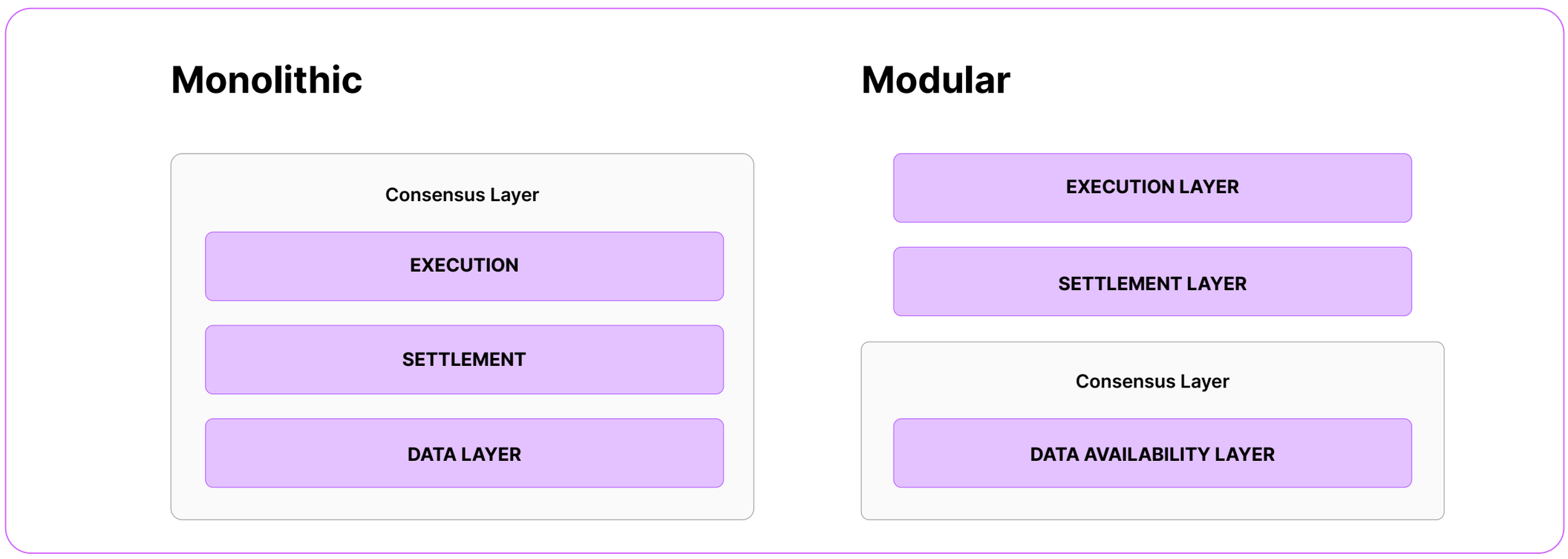
For developers, the benefits are clear: a flexible, pay-as-you-go model with global accessibility and transparent verification. For providers, the network offers a direct, permissionless path to monetise their hardware. In combination, these features turn compute into an open, verifiable, and economically sustainable resource, bringing AI closer to the decentralised ethos that 0G represents.

Consensus and Security

Consensus is the backbone of any blockchain, and in 0G it takes on an even larger role: coordinating multiple specialised networks for storage, data availability, and compute into a single operating system.

The challenge is to support AI-scale workloads without fragmenting security or creating bottlenecks. Running AI processes on traditional blockchains illustrates the problem clearly. Executing even a simple model on Ethereum would cost millions in gas fees, throughput is capped at around 15 transactions per second, and the chain cannot natively handle the data volumes that AI requires. To overcome these limitations, consensus in 0G has been designed for parallelism, modularity, and shared security.

MONOLITHIC vs. MODULAR BLOCKCHAINS



Source: Celestia

At its core, the 0G system runs many parallel consensus networks, each responsible for a subset of activity. These networks are not isolated: they are bound together by a shared staking model that ensures all subnets inherit the same level of security. In this way, consensus is not a single bottleneck but a horizontally scalable fabric that grows with demand.

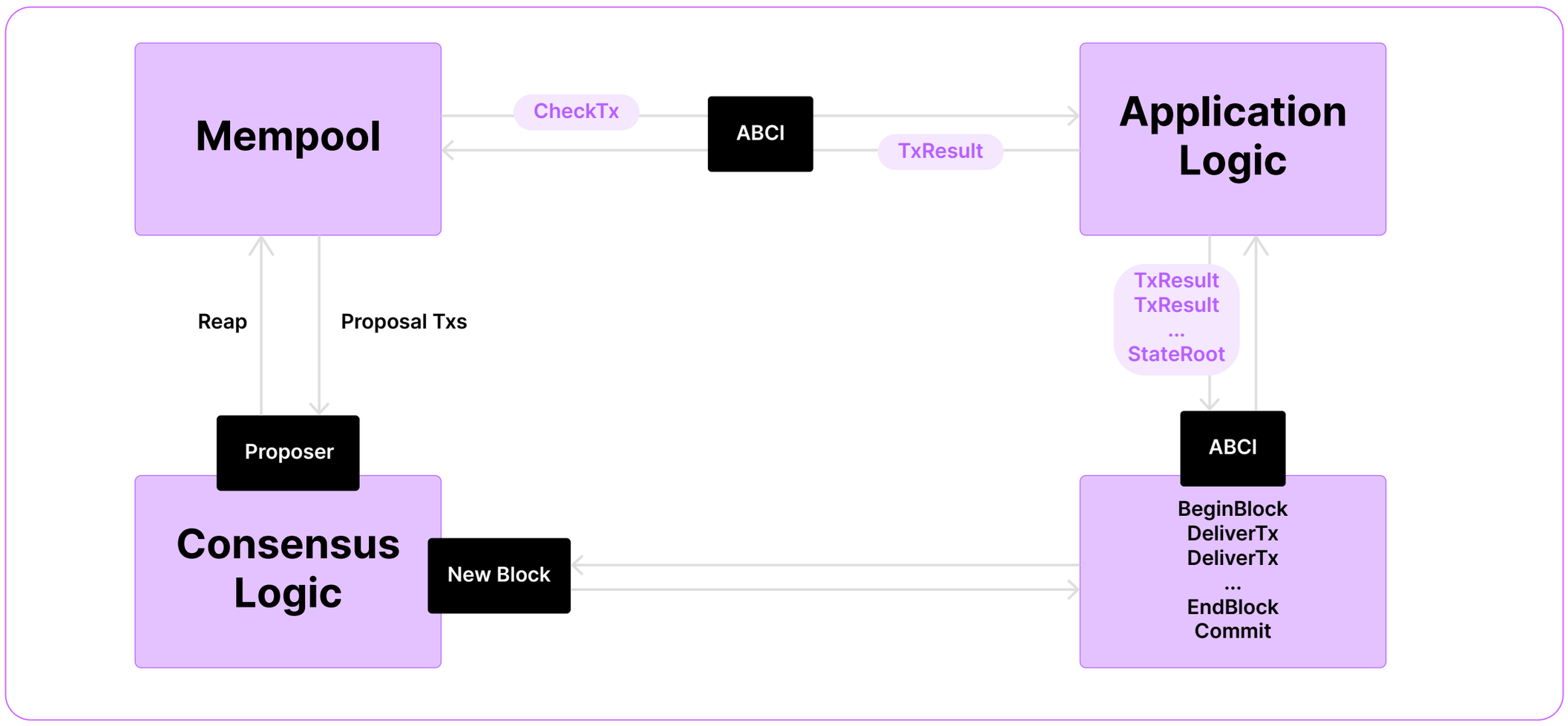
CometBFT and the Evolution of Consensus

Consensus is the mechanism that allows decentralised networks to agree on state. In 0G, it does more than simply finalise blocks; it acts as the coordination layer that binds storage, data availability, and compute into a single operating system. To understand why 0G relies on CometBFT, it helps to look at how consensus has evolved over time.

First launched in 2014 as Tendermint, it was one of the earliest Byzantine Fault Tolerant (BFT) protocols applied to blockchains. Its key breakthrough was proving that chains could achieve deterministic finality without the resource demands of proof of work. Rather than miners competing with raw computation, Tendermint relied on validators, where a two-thirds majority was enough to finalise a block. This made confirmations faster, timings more predictable, and the process far less energy-intensive.

The diagram below illustrates how this works in practice. Transactions enter the mempool and are first checked before being proposed for inclusion in a block. Validators coordinate through consensus logic, and once agreement is reached, the block is finalised. The Application Blockchain Interface (ABCI) links consensus to application logic, ensuring that the system's state is updated consistently and that both layers remain in sync.

COMETBFT MODEL



Source: CometBFT

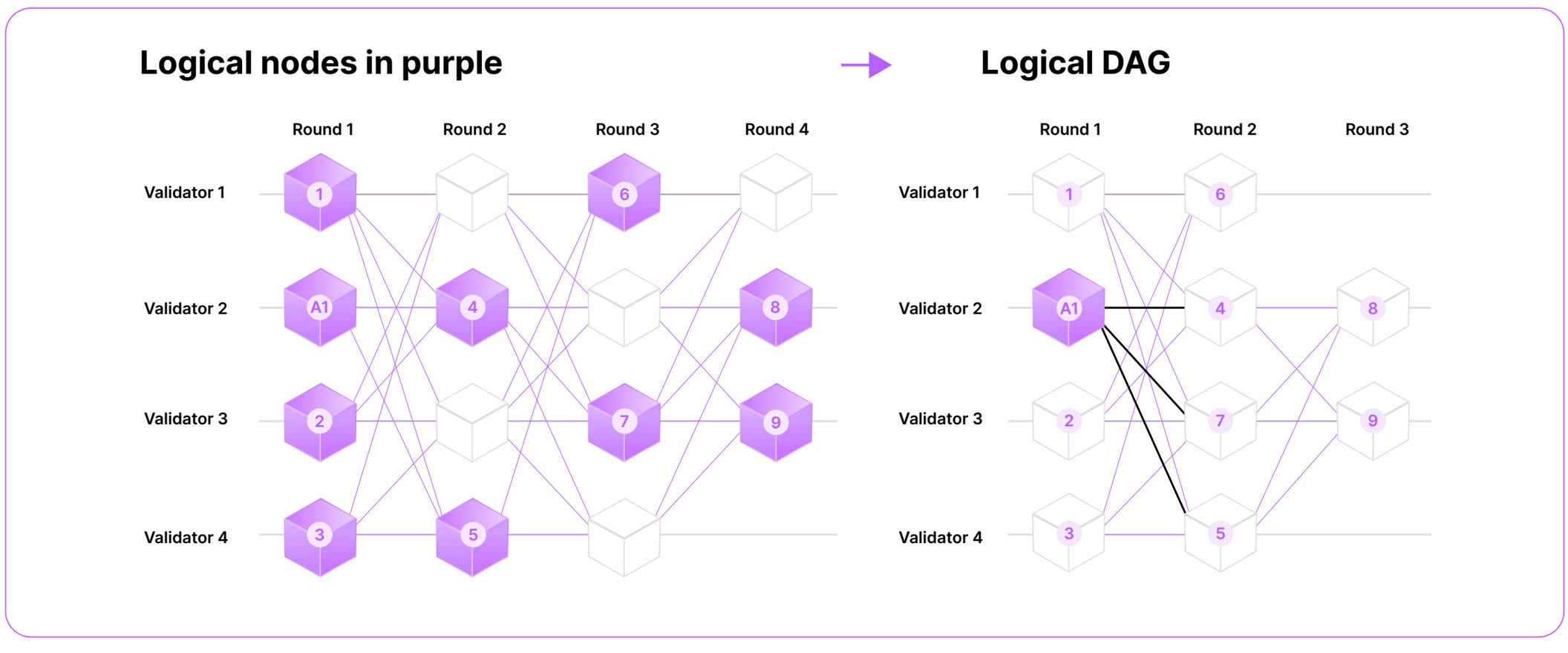
Over the years, this approach has proved remarkably robust. Tendermint became the backbone of the Cosmos ecosystem, securing billions of dollars across dozens of independent chains. In 2022, the protocol was renamed CometBFT, reflecting its evolution into a general-purpose BFT system, extending beyond Cosmos and setting the foundation for projects like 0G.

For 0G, the choice of CometBFT is deliberate. It avoids the risks of adopting untested consensus protocols and instead builds on one that has already been proven in live economic environments.

What 0G changes is not the core algorithm but the way it is configured. Block intervals and timeouts are tuned to favour throughput and responsiveness, allowing the system to reach more than 2,500 transactions per second with sub-second finality. These parameters are chosen with AI in mind: workloads that include continuous inference requests, rapid agent-to-agent communication, and frequent model checkpoints.

The roadmap extends beyond this. 0G plans to integrate DAG-based consensus, where multiple blocks can be confirmed in parallel rather than in sequence. This removes the bottleneck of linear block production.

LOGICAL DAG



Source: Decentralized Thoughts

For AI, where millions of micro-transactions may occur at the same time, DAG consensus could provide an order-of-magnitude increase in throughput. In effect, it takes consensus from thousands of transactions per second to potentially millions, aligning the system with the data intensity of AI training and inference

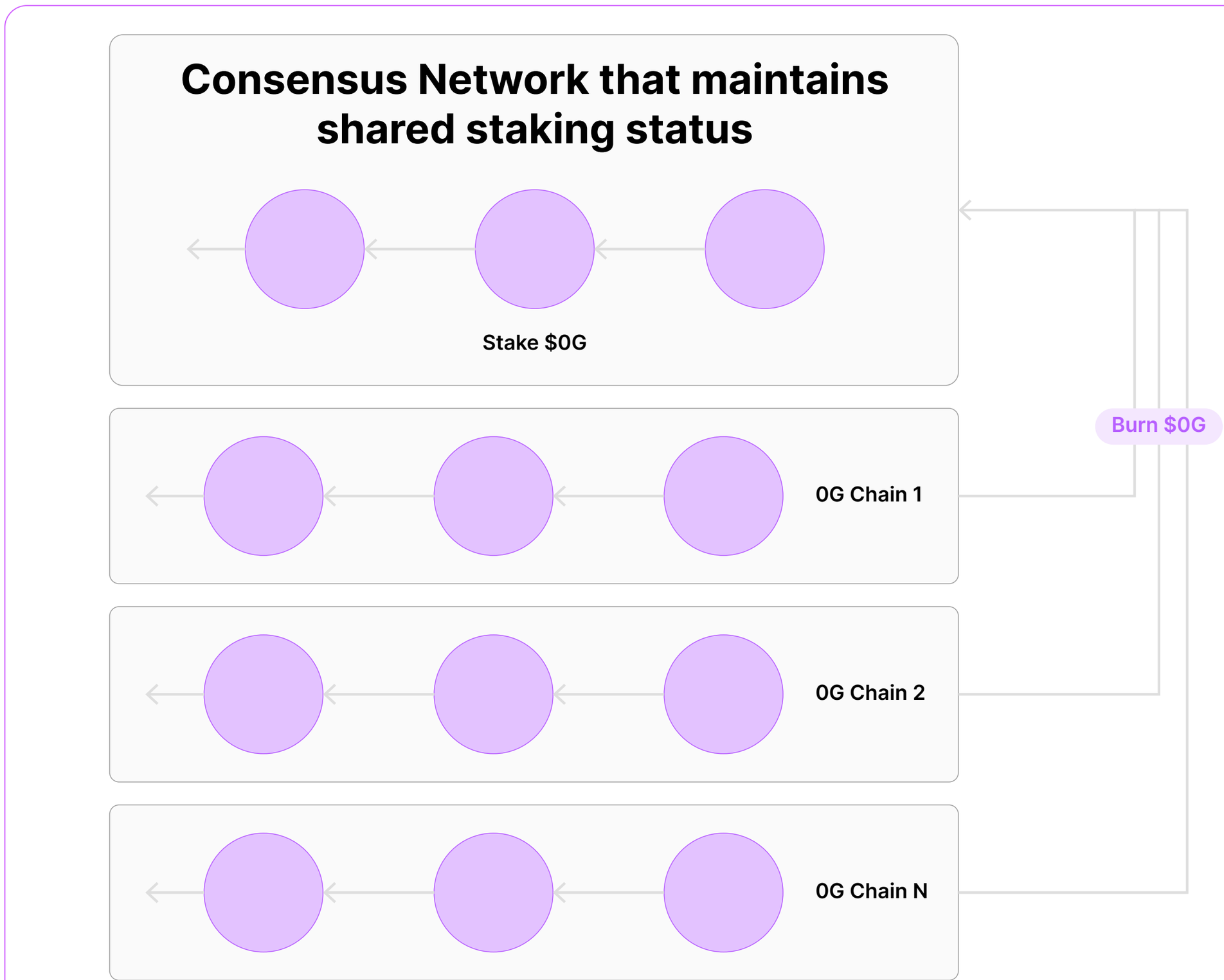
Multi-Consensus and Shared Staking

Alongside its consensus engine, 0G introduces a multi-consensus and shared staking model designed to solve two issues that affect modular blockchains: capital fragmentation and uneven security.

In most ecosystems, each new rollup or subnet must recruit its own validator set and collateral. This dilutes security and forces capital to be spread thin. Smaller chains end up more vulnerable, while validators must either specialise in one domain or split their stake across many, reducing efficiency.

0G addresses this with shared staking anchored to Ethereum. Validators lock a single ERC-20 token at the root layer, and that stake secures every subnet, whether for storage, data availability, or compute. If a validator misbehaves anywhere, it risks slashing at the root contract. This creates a uniform incentive to act honestly and ensures that all subnets inherit the same level of security.

THE LIGHTWEIGHT, SAMPLE-DRIVEN CONSENSUS APPROACH



Source: 0G

This design provides three key benefits. First, it eliminates weak spots by making security consistent across the ecosystem. Developers building on 0G do not have to weigh the relative safety of different subnets; all are equally protected. Second, it improves capital efficiency. Validators earn rewards from multiple services with one stake, putting their capital to better use. Third, it supports horizontal scaling. As demand grows, more consensus networks can be added without weakening the system or fragmenting validator participation.

The architecture is strengthened further by a modular separation of consensus and execution. Consensus upgrades, such as performance tweaks or security fixes, can happen independently of execution, while execution layers can adopt Ethereum innovations like account abstraction or EIP-4844 without changes to consensus. For AI developers, this means a system that evolves quickly while maintaining stability.

Economically, validators are compensated through block rewards, fees, and staking yields, while penalties for downtime or misbehaviour maintain discipline. Because a single stake covers many subnets, rewards scale with participation, encouraging validators to support the system broadly.

The result is a consensus fabric that is secure, modular, and aligned with AI's needs. Developers get Ethereum-level security without throughput bottlenecks. AI applications gain an integrated environment where storage, compute, and data availability are coordinated under one trust model. And end users benefit from faster, cheaper, and more reliable operations, backed by one of the most established consensus engines in blockchain history, adapted for decentralised intelligence.

Higher-Level Innovations

With the architectural foundations of 0G established, it is important to examine the innovations that operate above the core stack. These components are not strictly necessary for the system's functionality but are critical for its adoption and long-term relevance.

For developers, they lower the barrier to building trustworthy AI systems. For users, they strengthen guarantees around safety, ownership, and accessibility. Two features stand out: AI Alignment Nodes and DePIN Integration.

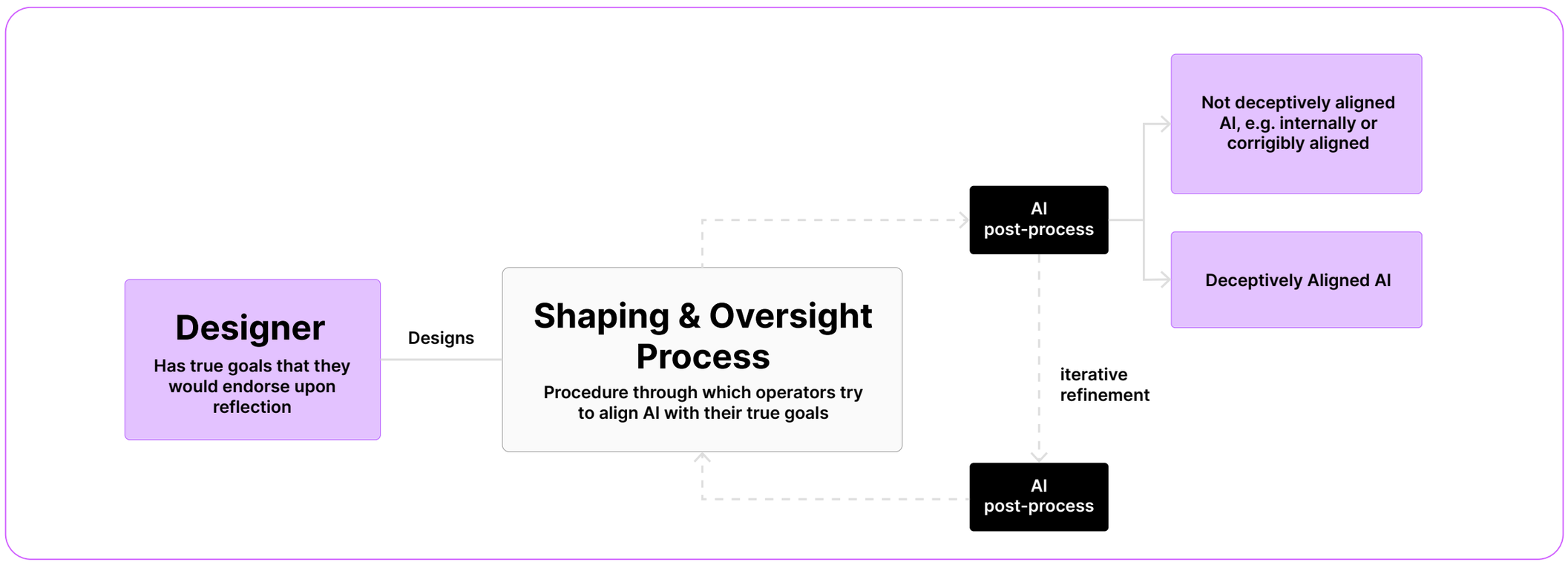
AI Alignment Nodes

A major challenge for decentralised AI is not only performance but alignment: ensuring that AI systems behave in ways that are reliable, transparent, and safe. In traditional blockchains, alignment stops at verifying transaction validity. But once intelligent agents are introduced, the question becomes whether their behaviour itself can be trusted.

Centralised platforms like OpenAI or Anthropic address this with internal monitoring and safeguards. While this works to an extent, it is neither transparent nor accountable, and it depends entirely on the incentives of a single organisation.

This is where the risk of deceptive alignment arises: an AI system may appear to act according to oversight during training or evaluation, but pursue hidden objectives once deployed. In a decentralised setting, alignment mechanisms must therefore be open, verifiable, and resistant to both errors and incentive drift.

PROCESS-ORIENTED VIEW OF DECEPTIVE ALIGNMENT



Source: Alignment Forum

0G's answer is AI Alignment Nodes. These are specialised participants whose role is to watch over both models and the network:

- Model monitoring. They track whether AI outputs remain within expected parameters, detect drift, and flag anomalies such as hallucinations or bias.
- Network oversight. They check for protocol violations or malicious activity, reporting problems back to governance.

What makes this design important is its integration into the protocol itself. The information Alignment Nodes gather can directly inform governance, giving the community evidence when deciding how to handle safety issues or update rules. This builds accountability into the infrastructure rather than leaving it as an afterthought.

For 0G, this creates a competitive advantage. Many blockchain-AI projects focus only on raw infrastructure, offering compute or storage but leaving alignment risks to developers. 0G, by contrast, embeds oversight at the network level.

For developers, this reduces the cost of building trustworthy AI applications. For users, it assures that models are actively monitored. And for regulators, it offers a framework that meets the growing demand for auditability and safety controls in AI systems.

DePIN Integration

Decentralised Physical Infrastructure Networks (DePIN) represent a new way of scaling hardware resources. Instead of relying on centralised cloud providers that build and operate data centres, DePIN models pool underutilised hardware from around the world and coordinate it with blockchain-based incentives. This approach lowers costs, broadens geographic reach, and reduces reliance on single providers.

For 0G, DePIN integration is a natural extension of its design. AI workloads require access to specialised hardware, particularly GPUs, and traditional options are expensive and concentrated in the hands of a few firms.

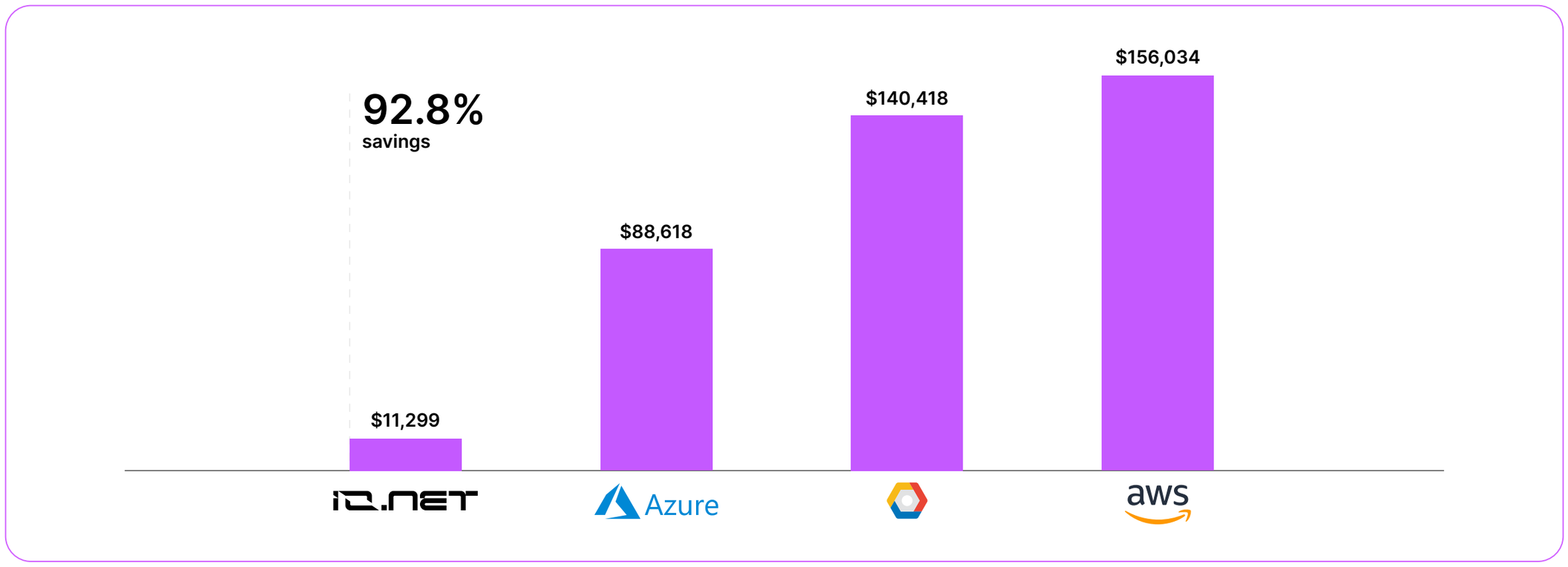
By tapping into decentralised GPU networks, 0G Compute can provide the scale needed for training, inference, and application-level AI services without replicating the capital-intensive model of cloud infrastructure. Instead, it harnesses distributed resources and brings them into its operating system for decentralised AI.

The benefits are clear. DePIN networks offer cost efficiency by making use of hardware that would otherwise sit idle, often reducing costs by more than half compared to conventional providers.

They are also geographically distributed, which improves latency and allows developers to access compute capacity closer to their end users. The distributed design increases resilience by avoiding single points of failure, while scalability comes from the ability to onboard new providers into the network without physical expansion.

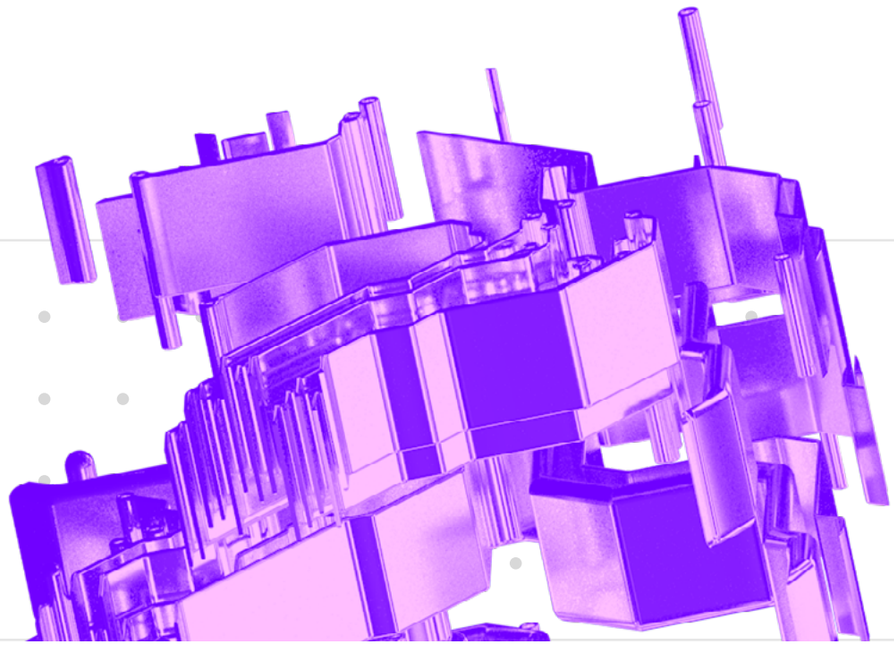
Partners and Ecosystem

0G has already aligned with leading DePIN projects. io.net operates one of the largest decentralised GPU networks, with more than 300,000 verified GPUs contributed by data centres, miners, and individual operators across over 130 countries. Its infrastructure can deploy GPU clusters in under two minutes and offers substantial cost savings compared to cloud incumbents.

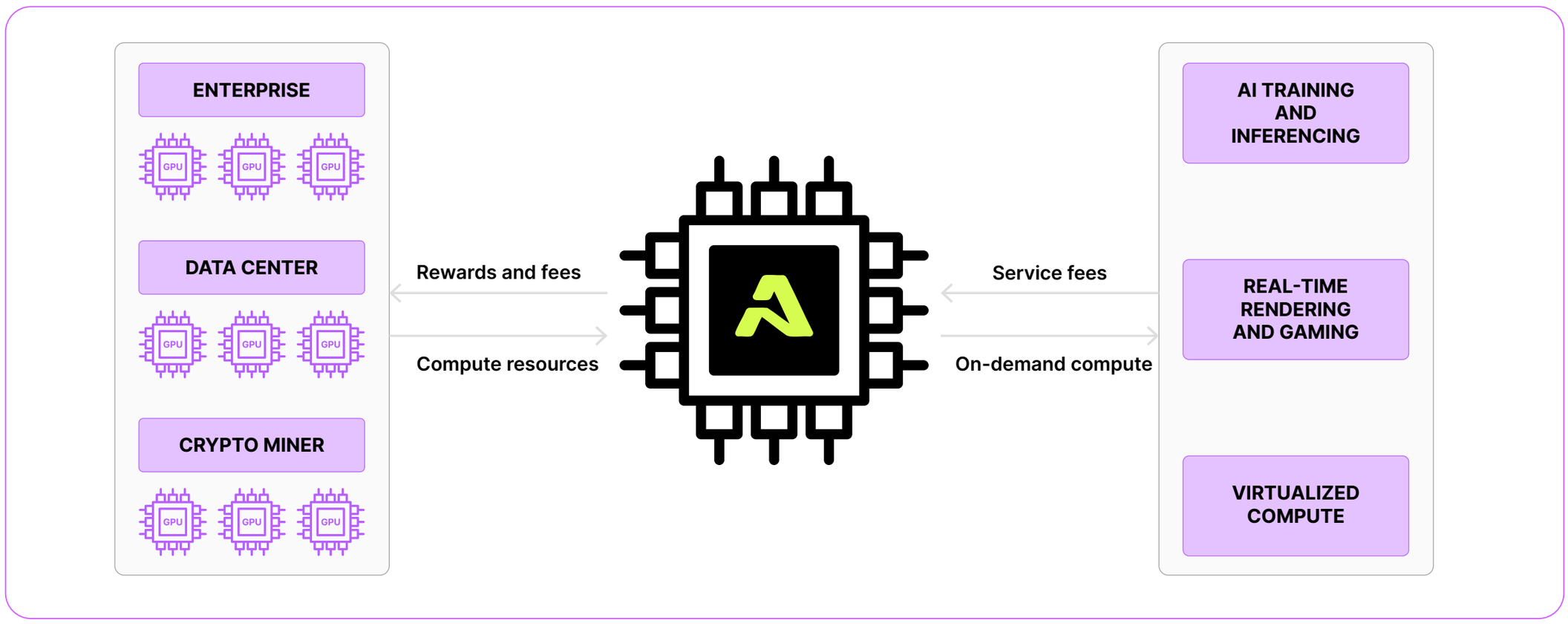


Source: io.net

Aethir brings a complementary focus on enterprise-grade GPU-as-a-Service, with more than 40,000 high-end GPUs and infrastructure designed to meet the demands of AI, gaming, and virtualisation, supported by uptime guarantees and verification mechanisms such as Proof of Rendering.



AETHIR'S DECENTRALIZED GPU NETWORK



Source: Aethir

Together, these integrations position 0G to meet the demand for decentralised AI computing at global scale. Developers gain access to affordable, verifiable GPU power without the barriers of centralised providers. Enterprises can deploy workloads on infrastructure that is both cost-effective and resilient. For users, it means AI services can be built and delivered on a foundation that is transparent, decentralised, and broadly accessible.

By combining DePIN infrastructure with its own consensus, storage, and compute layers, 0G creates a vertically integrated stack that can support advanced AI applications from end to end. This strengthens the ecosystem's attractiveness to developers who want reliable access to compute, while also making 0G one of the few platforms capable of connecting decentralised AI software with decentralised hardware at scale.

Competitive Landscape

0G is best understood not as another single-purpose chain but as an operating system for decentralised AI that combines data availability, storage, and compute into one coordinated stack. Most alternatives specialise in only one of these layers, which makes apples-to-apples comparisons tricky.

To compare it fairly, we look at the three places builders actually decide between today. First, general-purpose blockchains where applications live, which are Ethereum with its rollups and Solana. Second, dedicated data availability layers, which are Celestia and EigenDA. Third, storage protocols, which are Filecoin, Arweave, and IPFS.

The question in each case is how well the system handles AI realities, large immutable datasets, fast mutable state, high read and write rates, verifiability at the protocol layer, and clean integration across the stack.

Blockchains

Blockchains are where most AI projects first deploy, but they were not designed for AI's bandwidth and storage demands. They fall into two families: monolithic chains such as Solana, and modular ecosystems centred on Ethereum and its rollups. Both families have strengths, but they also externalise the heaviest data flows, leaving AI teams to assemble additional infrastructure.

Solana vs 0G

Solana achieves very high throughput with Proof of History and pipelined validation. This makes it efficient for transaction-heavy consumer applications. The trade-off is that all activity shares the same ledger. There is no separation between DA, storage, and execution, so large datasets and mutable state must live outside the chain. For AI pipelines, this means raw execution is fast, but storage and data verification still require external systems.

0G splits these layers explicitly. Consensus remains lightweight while data availability and storage have their own protocols and proofs. The result is that AI developers get the same speed advantages for coordination but without pushing their heaviest workloads off-chain. Where Solana is a single ledger with impressive speed, 0G is a horizontally scalable fabric designed to carry AI-scale data.

Ethereum vs 0G

Ethereum anchors the deepest liquidity and strongest security, but it was never built to handle large datasets. Storage on Ethereum is prohibitively expensive, and archival data is offloaded to third parties. Even with EIP-4844, data blobs are only temporary. For AI applications, this creates fragmentation: training data, logs, and indices must be stored elsewhere, breaking provenance guarantees.

0G integrates storage and DA as native services. Immutable logs and mutable key-value data are verified by Proof of Random Access and secured under the same consensus root. This allows AI builders to treat training data and state as part of the chain rather than as external dependencies. Ethereum ensures contracts settle credibly; 0G ensures data itself is credible.

L2s vs 0G

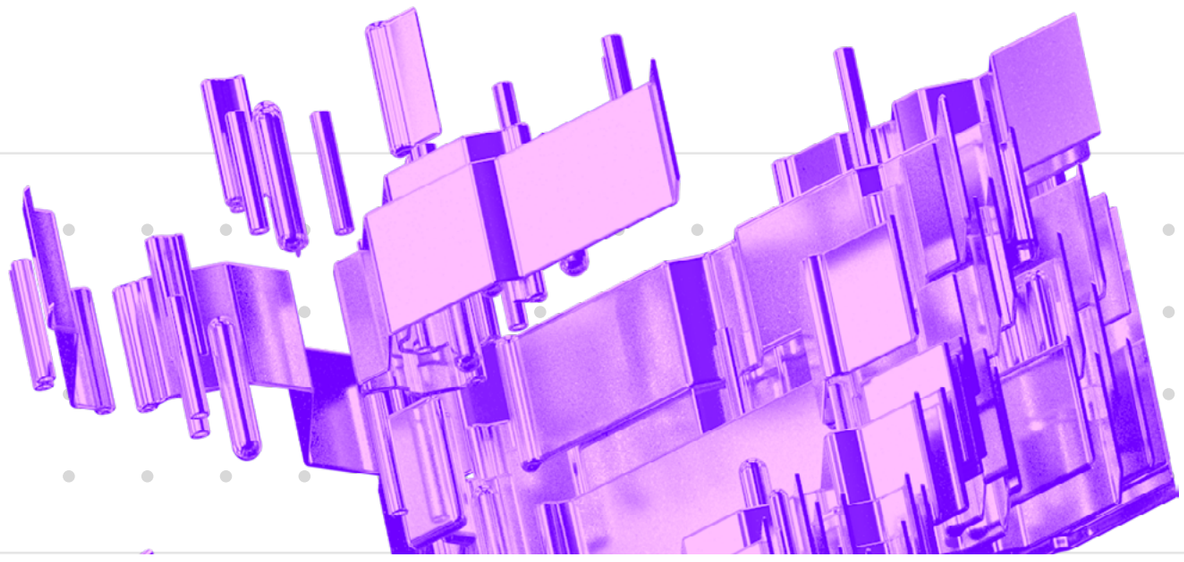
Ethereum rollups improve throughput and costs, but they also fragment execution. Each rollup must secure its own DA and storage setup, often choosing Celestia or EigenDA. The result is an ecosystem of many chains, each with partial guarantees. For AI builders this adds complexity, since storage and compute must be replicated per rollup.

0G avoids this by using shared staking across all its specialised networks. Storage, DA, and compute are coordinated under the same security model. Instead of hundreds of small ecosystems, AI developers work inside one integrated operating system where guarantees are uniform.

Conclusion

Blockchains provide strong settlement or throughput, but none were designed as data systems for AI. Solana excels in speed, Ethereum and its L2s excel in security and composability, yet both externalise storage and DA. 0G combines their strengths while embedding data availability and storage directly into its architecture.

FEATURE	SOLANA	ETHEREUM	L2S	0G
Throughput	Very high	Low on mainnet	Higher on rollups	Horizontally scalable across subnets
Storage	Limited, off-chain	Very costly, archival off-chain	Externalised	Integrated log + KV storage
DA	Not separate	Blobs, temporary	Outsourced to DA layers	Native DA integrated with storage
Security	PoH validators	Strongest L1 settlement	Tied to Ethereum	Shared staking with root validators
Fit for AI	Fast execution but no storage	Secure settlement but fragmented data	Fragmented across many rollups	End-to-end system for storage, DA, compute



Data Availability Layers

Data availability ensures that published data can be retrieved and verified, which is essential for rollups and modular blockchains. The leading examples are Celestia and EigenDA. Both have advanced the state of DA, but neither was designed with AI's needs for long-lived, mutable datasets.

Celestia vs 0G

Celestia pioneered data availability sampling, making it possible for light clients to verify large blocks efficiently. This works well for rollups that only need to confirm publication. But Celestia does not integrate storage. Once data has been published, persistence and fast access are the developer's problem. For AI pipelines, this creates gaps, since models need to repeatedly retrieve and update data.

0G integrates DA with storage. Availability checks are performed directly on data that lives in the storage layer, so persistence and verifiability come from the same protocol. This makes retrieval as reliable as publication.

EigenDA vs 0G

EigenDA scales throughput by tying availability to Ethereum restaking and operator committees. This alignment with Ethereum is valuable for rollups, but EigenDA is not optimised for workloads outside that scope. Like Celestia, it does not handle mutable state or storage directly, forcing AI developers to build on multiple systems.

0G removes that burden by combining erasure-coded DA with its log and key-value storage. Availability, persistence, and retrieval are guaranteed within the same framework, secured by shared staking.

Conclusion

DA layers have expanded Ethereum's scalability, but their scope remains narrow: they guarantee that data is available, not that it is stored, queryable, or mutable. 0G extends DA into a data plane that is tailored for AI.

FEATURE	CELESTIA	EIGENDA	0G
Core design	DAS with erasure coding	Restaking with operator committees	DA + storage integration
Storage coupling	External only	External only	Native storage and KV
Security	Sampling, light clients	Ethereum validator alignment	Shared staking and VRF-selected quorums
Fit for AI	Confirms publication, no persistence	High throughput, no storage	Publication + persistence + retrieval

Storage Protocols

Storage protocols preserve data, but their assumptions differ from what AI requires. Filecoin focuses on replicated persistence, Arweave on permanence, and IPFS on addressing. All three are important, yet none combine permanence with low-latency mutability and verifiable retrieval.

Filecoin vs 0G

Filecoin's Proof of Replication and Proof of Spacetime make it reliable for archival storage. The weakness is in retrieval and updates. Latency depends on retrieval miners, and there is no native key-value interface for mutable state.

0G addresses this with its two-layer design: an immutable log for archival data and a key-value runtime for mutable workloads. Proof of Random Access ensures that providers can retrieve chunks quickly, rewarding useful I/O rather than raw capacity.

Arweave vs 0G

Arweave is designed for permanence, using a blockweave and endowment model to ensure data lasts forever. It is ideal for static artefacts but not for dynamic datasets that change constantly during training or inference.

0G covers both. Permanent datasets live in the log, while mutable state lives in the key-value runtime. Both are bound by verifiable proofs of storage and availability.

IPFS vs 0G

IPFS provides efficient content addressing and distribution but leaves persistence and incentives to external pinning markets. For AI teams, this means guarantees depend on additional arrangements.

0G integrates addressing, persistence, and incentives natively. Developers know that their data is not just addressable but also provably stored and retrievable.

Conclusion

Filecoin, Arweave, and IPFS each solve part of the storage puzzle. 0G combines them into one design that supports permanence, mutability, and verifiability under a single security model.

FEATURE	FILECOIN	ARWEAVE	IPFS	0G
Persistence model	Replication and spacetime proofs	Permanent blockweave	Content addressing, no guarantees	Log + KV with endowment
Mutability	Limited	None	Partial via IPNS	Full KV runtime
Proof system	PoRep, PoSt	Endowment-backed permanence	None native	Proof of Random Access
Retrieval	Market latency	Gateways	Peer-based	Fast, verifiable
Fit for AI	Archival only	Immutable only	Transport only	End-to-end archival + mutable state

Overall Assessment

The comparisons make one pattern clear. Existing blockchains, DA layers, and storage protocols all deliver important advances, but none were designed with AI as their central use case. Ethereum and Solana provide settlement and execution but push storage and data verification elsewhere. Celestia and EigenDA offer strong availability but no persistence or mutable state. Filecoin, Arweave, and IPFS secure data but cannot deliver the low-latency mutability or unified proofs that AI requires.

0G succeeds because it integrates these fragmented capabilities into one operating system. Storage, availability, and compute are not separate add-ons but parts of the same verifiable data plane. For AI developers this reduces complexity, lowers coordination risk, and ensures that the guarantees they rely on are embedded at the protocol layer. The result is a system that does not just compete with each category but consolidates their strengths while eliminating their weaknesses. That is why, when you put everything on a single scorecard, 0G clearly comes out as the winner.

COMPETITIVE SCORECARD

CATEGORY / FEATURE	ETHEREUM & L2S	SOLANA	CELESTIA	EIGENDA	FILECOIN	ARWEAVE	IPFS	0G
High throughput for coordination	✗	✓	✗	✓	✗	✗	✗	✓
Strong settlement security	✓	✓	✗	✓	✗	✗	✗	✓
Native data availability	Partial (blobs only)	✗	✓	✓	✗	✗	✗	✓
Integrated storage	✗	✗	✗	✗	✓	✓	✗	✓
Mutable KV runtime	✗	✗	✗	✗	✗	✗	Partial	✓
AI-fit architecture	✗	✗	✗	✗	✗	✗	✗	✓

Risks & Challenges

Like any blockchain or crypto-native infrastructure project, 0G carries risks that need to be understood alongside its strengths. Its design is ambitious and its roadmap extends across multiple domains, from storage and data availability to compute and consensus. While the architecture addresses critical bottlenecks in AI infrastructure, execution and adoption will determine its long-term viability. Five areas stand out as central challenges.

1. Technical Complexity

0G's design integrates four specialised layers: storage, DA, compute, and consensus into one operating system. Coordinating these modules smoothly at scale requires not just protocol engineering but careful optimisation of performance trade-offs. Bottlenecks in one layer, such as DA sampling throughput or storage replication, could affect the overall system. Even with modular separation, the challenge of integration at AI scale remains significant.

2. Ecosystem Adoption

Building a new ecosystem from scratch is difficult in an environment where Ethereum, Solana, and other established chains already offer liquidity, developer mindshare, and tooling. For 0G, success will depend on attracting early developers with clear advantages and producing flagship applications that prove its superiority for AI workloads. Without strong onboarding and integrations, even technically superior infrastructure may face slow adoption. So far, however, 0G has shown strength on this front, launching with more than 100 partners during its mainnet release.

3. Sustainable Incentives

The incentive model underpins the health of the network. Proof of Random Access, DA rewards, and the compute marketplace must balance profitability for operators with affordability for users. If small providers cannot compete effectively, decentralisation may suffer. Conversely, if rewards outstrip demand, inflation could undermine sustainability. Designing incentives that adapt as the ecosystem grows is a key challenge.

4. Competitive Environment

0G does not compete in isolation. Other decentralised protocols are advancing in DA, storage, and compute, while hyperscalers such as AWS, Azure, and Google are rapidly expanding AI-optimised infrastructure. 0G's edge lies in offering verifiability and decentralisation, but it must continue to differentiate on performance and cost efficiency to stay relevant in both the crypto-native and enterprise markets.

5. Regulatory Uncertainty

As AI and blockchain converge, regulatory scrutiny is likely to increase. Questions around data ownership, compliance for decentralised compute marketplaces, and jurisdictional differences in storage and privacy laws may all impact 0G's adoption. A decentralised system cannot eliminate these risks, but it must design governance and compliance pathways that reassure institutional users without undermining its core principles.

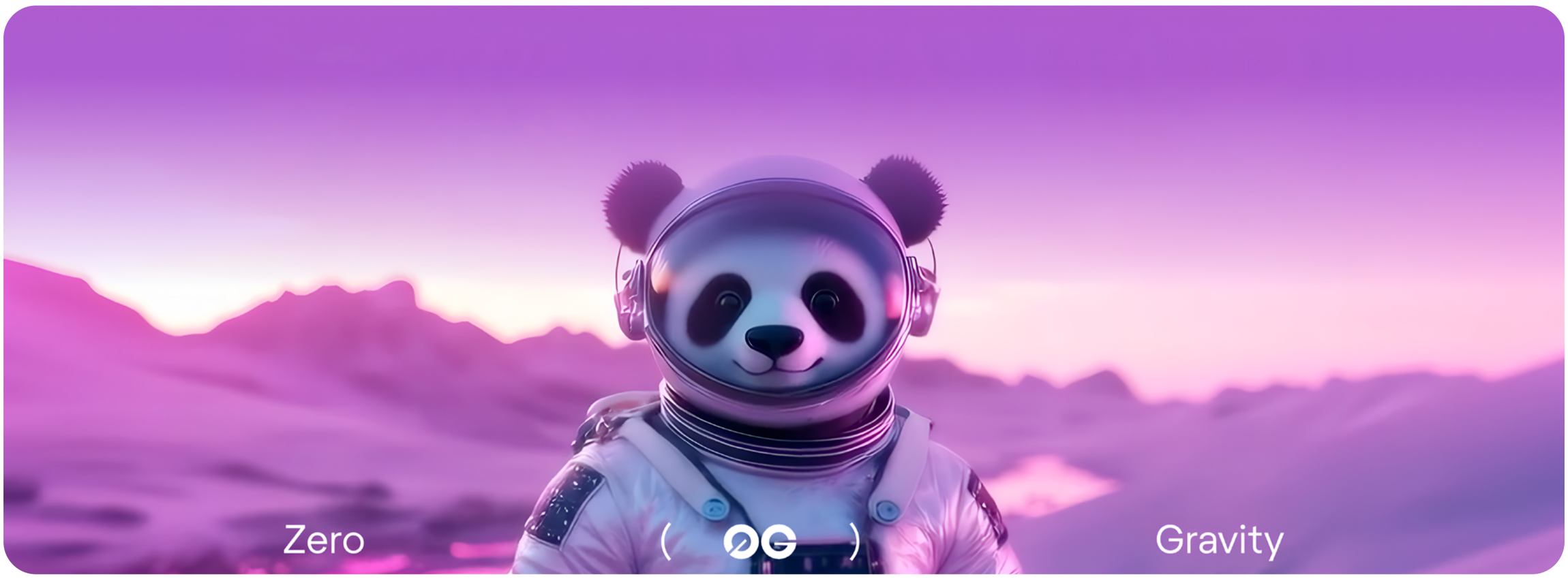
Outlook and Conclusion

AI is on track to become the defining industry of the coming decade, with trillions of dollars flowing into compute, storage, and infrastructure. The question is not whether demand will grow, but who will control the systems that deliver it. Today, a handful of hyperscalers set the terms of access, keeping AI closed, costly, and opaque. The opportunity is to build an alternative foundation that is decentralised, transparent, and verifiable by design.

0G is positioned to be that foundation. By integrating storage, data availability, and compute into a single operating system, it removes the fragmentation that has forced developers to stitch together multiple protocols. Instead, it offers a unified environment built specifically for AI workloads. Data can be stored immutably or updated in real time, availability is guaranteed and provable, and compute is delivered through an open marketplace where outputs can be verified. This is not a narrow optimisation of one layer but a systemic solution for AI-scale infrastructure.

The risks are clear. Adoption will take time, incentives must remain sustainable, and competition from both decentralised protocols and cloud incumbents will be intense. Yet the architecture and timing give 0G a strong position. The launch of the Aristotle Mainnet, with over one hundred partners at day one, shows that the ecosystem is not starting from zero but already mobilising at scale.

In the long term, the opportunity for 0G is to become the backbone of decentralised AI: the place where data is stored and proven, where compute is accessed and verified, and where applications can run without reliance on corporate silos. If it succeeds, 0G will not just be another blockchain in the modular stack. It will be the operating system that makes AI open, auditable, and accessible as a global public good.



DLResearch × 0G

Building the Foundation for an Open AI Economy: The Case for 0G