The InferProof Essay

by: Nick Quick

August 14, 2026

The InferProof One Bluepaper does a fantastic job of laying out the conceptual design of the project. It covers all the bases, gives prospective miners the data they'll need to make a Minecart decision, and outlines the various security protocols and responses that are formalized. What it doesn't do is hammer the why. That's what this entirely human written essay will do. It will also leverage crypto vernacular, trenches slang and inevitably cast critiques over the last ten years of retail crypto behavior.

If we travel back a decade or so, we arrive at Bitcoin & cryptocurrency's first dance with early adopter retail. The "one cpu one vote" vision drifted away as ASICs took over BTC's network. In turn retail looked at alts - non-BTC crypto assets that ran on their own networks. Alts were mineable with consumer grade GPU/CPU. ETH was $11.23 on 14 Aug 2016, mineable, and offered something new via smart contracts. ERC-20s themselves weren't mineable, but since they didn't have to worry about network consensus they bloomed into tokens with immense and diverse use cases. Monero was (and still is) a beautiful concept that spawned a variety of forks, each running their own chains with their own mining software and forked CLI wallets. The author was a core team contributor to one of these Monero forks and participated heavily in the CryptoNight mining scene.

At that moment use cases were generally just "currency," at least for everything not ERC-20. Each new currency was meant to be spent, and each new network competed against each other for hash rates. If 10 new coins launched L1s everyday on BTC Talk via [ANN]s, then those 10 had to attract miners away from the other 500+ mineable L1s floating around in 2016 - 2017. A PoW chain without any hash power has no security and is DOA. Announcing, recruiting, and deploying software for miners & traders - this was all a laborious process that went through the meat grinder of social validation on BTC Talk. Scams were far

less present because bootstrapping a new coin/token required legwork. The main question remained for all new launches though: how to attract miners?

Well like anything - incentives. Gamified emission schedules and bootstrapping real liquidity on a semi-decent exchange were blue chip prospects. For the miners in the PoW trenches of yesteryear, stories about sketchy exchanges in all Mandarin, where they're navigated by domain slug and random clicks, exist. Exchange closures, collapses and hacks were morning reads on the Twitter feed. Many will say In-N-Out is a burger joint, but these miners will say In-N-Out was the only viable exchange strategy that existed. If a new PoW chain was launching, and they were able to secure a listing on an exchange written in English, and their emissions meant early miners were going to eat, they'd win. For at least a few weeks, until the emission schedule slows, the miners leave, the hype dies, and the traders chase the next 2x over a 4-week period. Because back then a 2x in a short period of time was a remarkable return. Retail pivoted to unrealistic expectations with the rollout of launchpads, prediction markets, liquidity pools, and similar products. It wasn't always pure degeneracy. Miners back then were willing to lend their real, tangible resources to the network for a potential return, and the networks were happy to receive it. Far more symbiotic than adversarial at root.

Yet this model of proof-of-work gamification was wholly unsustainable, and the market abandoned it at the base layer. PoS (Proof-of-Stake) networks came up. Protocols like Ark and Neo offered returns for "security & consensus" by keeping the coin balance in wallet rather than on an exchange. "Passive income" and "Eco-friendly consensus" were all the rage as PoS further developed. PoS was touted as more secure, because PoW networks were routinely re-org'd, 51%'d, chain-forked, or simply shutdown. People would mine on a chain, hold the asset, step away, then come back and it's all worthless. Many readers have experienced this without the mining element, but the PoW-mining relationship is dramatically different than any staking network's is.

In a PoS network I expend money (via market buying the coin) for network security and consensus. In turn I'm offered a return by the protocol, stake pools, consensus validator, coordinator, or whomever. The return is

generally trickle-down, low APR and insider fueled. Consensus, returns, protocol changes are all beyond anyone's personal ability to affect. The singular network will always be just that (unless it grinds to a halt), and retail users will take what is given.

In a PoW network I expend money on energy and hardware to secure a decentralized network. I can do this by myself, or in a pool with other miners and earn the entire or part of a block reward. This cycle continues for as long as the chain is viable and operational, according to the protocol's rules. If I or a group of miners disagree with protocol changes, we can maintain the original chain and not observe the fork, thereby generating two chains out of one. See Ethereum Classic vs. Ethereum. In a PoW model retail users can steer direction with hash power. Unfortunately, in modern times high end GPUs or specialized mining equipment is out of reach for most of the globe and the work resolves to who has the most capital to throw at hashing - not how distributed the power is. So the west & highly developed east are the only players that remain.

Satoshi's Whitepaper says nothing about massive mining farms using nation state power levels to secure the chain. It says the opposite - one cpu one vote. Bitcoin is remarkable and the crown jewel of anything crypto. It's the hardest nut to crack and it has stood the test of time by shirking the conceptual design for market realities in production. Yet Monero has been the PoW network to uphold decentralization above all else in its approach to mining. ASIC resistant, CPU focused, randomization. Even so, Monero has suffered attacks and had entities gain 51% of its network. If after 10+ years the concept of (real) decentralized security is nearly impossible for those two assets to attain, then any claim on doing it fresh in 2026+ should be met with supreme skepticism.

As the founder of an AI centric company, stumbling upon PoW concepts that incorporated LLM inference was immediately exciting. Perhaps this offered a different way to derive consensus via the input/output cycle that is going to dominate the future of business and consumer behavior. Truly the idea is wonderful. Yet, after a month or so of observation, it became clear the same PoW pitfalls are ever-present. Further, the invocation of inference into the PoW model allows for a variety of new, exotic attack vectors like fake/unvalidated

hash submissions. The Bluepaper contains further pointed references on this topic and real-world examples from Summer 2026 validate this suspicion.

Nevertheless, what stood out was that within these inference PoW models, the driver of the economic activity - the buyer of inference - is left with nothing but the LLM output. Sure, that is what they paid for and received, but any advanced user of AI will tell you not every output is the same. Models hallucinate, bloviate, dictate, lie, cheat, get mad, etc. The assumption that every output the inference buyer receives provides adequate value for what they paid is wrong. Very wrong. Removing this buyer from the coin mining process is illogical.

The buyer of inference is the person solely responsible for the economic levers stimulating inference PoW chains. The inference buyer provides demand and capital, each with unique wants/needs for their request. Conversely the miners provide the inference, but FOSS LLM hosting is a commodity. Nearly every major open weight model is available in some form or another across Fireworks, Hugging Face, Open Router, etc. Retraining an open weights model, forcing inference buyers to use “your model” to regulate PoW hashes is a captive, not open, system. There are trillions in weekly token demand for Deepseek v4 in all its forms, while there is legitimately zero native demand for Deepseek v<*PoW L1 retrained*>. The only reason people would specifically seek to use a model such as this is to farm tokens, but if the person buying the inference receives none…they have no reason to use it in the first place.

The question remains: how on earth does the buyer get left out of an inference mining process? Any coin or token generated from the product of inference should return to the inference buyer in some way, not explicitly to the provider. The provider sets the rate to use inference and earns a profit off that. Now, within a run of the mill inference PoW system, the entity running $250,000 worth of Blackwells gets a little perk on top of the direct revenue via coin mining? Nah. This review of a general inference PoW structure, not a specific coin’s implementation, is what led to the creation of InferProof One’s model:

The inference buyer is the miner.

If the live systems are inverted, and the miner is now the inference buyer instead of provider, the new apparatus must be appealing to prospective buyers/miners. The inference buyer is choosing a model that fits their needs, so the mining part must correlate and make sense. The question then is, what makes the miner tick? From observation and participation, these are key considerations for any PoW miner:

- *Energy* - how much energy is expended in the production of the PoW coin? This is generally denoted by the electricity rate miners pay to run hardware
- *Hardware* - what hardware is required to mine? In purest form the simplest hardware is the best. It has the widest distribution globally and allows for real decentralization. In most forms, it's a consumer or higher-grade CPU or GPU (rig).
- *Return* - what return can be generated by securing the network? This is the coin emission from each block and results in near instant liquidation at market price to cover operating costs. Mining calculators propose expected earnings, and if they outweigh energy costs you have a profitable mine. Distribution assumes buyers of these coins exist and can outweigh the dumps from miners. Reality usually says otherwise.

Of these three considerations, only one has existed within the realm of currency for all its history - energy. When we look at the energy consumed in inference, it is substantial. All of the energy used in inference is burnt and spent through the output. The inference as PoW models try to channel this but the inclusion of network security and consensus reduces the coin output to a prop. If a coin is meant to be mined during the product of real inference, yet millions are mined with fake outputs and fake hashes, where does that leave it? Are these products of exploitation forever in circulation pretending to be real outputs of inference, and the expectation is that the market just turns a blind eye to their true origin?

The goal of harnessing the time, money and energy spent on inference via portable blockchain tokens is worthy of pursuit. What's not necessary is to include block producing consensus within the formula. There's zero reason to tie

inference to consensus or security, and in the limited examples shown to date it's been disastrous - emergency soft forks, hard forks, rollback talks, exchange seizures, no communication, incompatible wallets, and losses for everyone along the way. Then it is all packaged and sold as useful or valuable work.

IP1's approach to these issues to remove all the noise. No consensus. No decentralization for the apparatus. No GPU or RunPod mining. No mysterious "South American datacenters" who run 70% of the network. Legitimately, the driver of economic activity in inference is the transaction between LLM provider and input/output end user. Everything external to that is noise. Turning inference into a portable blockchain object is entirely safe and possible if decentralized layer-one theater is thrown to the curb, which is what InferProof One has done.

The first challenge in this design was metering. Here were the options & considerations:

- Proxy API - miners register, generate an API key, pay through IP1. This model sucks, opens a ton of risk for Quick AI LLC, and requires payment processing for inference. Hard no.
- Oauth - OpenRouter oauth was the original conceptual design with complimentary vendors. When tested two major issues were discovered. First, the oauth produced a spendable key and Quick AI doesn't want that liability if exploited. Second, the oauth did not produce cumulative metering data. If miners used oauth, then IP1 would have to query token usage for each request and compile that way. At scale, this is not possible without inviting consistent tabulation errors.
- Administrative Key - these are offered by vendors like Anthropic and OpenRouter as non-spendable metering keys. OpenRouter calls their version Management Keys and Anthropic gates access to business orgs only. Admin keys report inference usage and dollars spent per model on a daily roll which is referenced within IP1. The roll turns at 0:00UTC each day and reflects what is billed in that period.

The only viable option for the concept was to use Administrative Keys. The primary metering goal of the apparatus is to verify legitimate, paid LLM usage. The above options provide two realistic ways to achieve this metering. One route requires serious formal commitment from miners by providing KYC/cc details to IP1, and the other is copy/pasting a non-spendable key. Choosing the Management Key model was one of the easiest conceptual decisions of the project after probing the realities of all options.

The essay's author was a believer in the ethos of crypto 10 years ago. Permissionless, decentralized, private, formal/capped emissions as anti-inflationary measures, ease, and global accessibility were the appeals. These were the features that drew early adopters, only to then later be largely abandoned for the never-ending pursuit of gains. In constructing IP1, the author wanted to channel the original ethos as best as possible, while still providing controls & regulation that come with a formal business operating the mining apparatus instead of an anon account on VPN. Despite the deployment from a formal business, IP1 carries the purest launch possible. No team allocation, no premine, no dev cut of each mined block. IP1 operators will start mining on Sept. 1 the same as anyone else, and mine only according to their paid usage; which is expected to be about ¼ of the maximum mineable amount per account.

In the end IP1 does not capture your login credentials (oauth login bound to pID), your spendable API keys (mgmt/admin key only), your credit card (on-chain payments) nor gate entry (any OpenRouter user eligible). The app itself is permissionless to enter and mine. Payments and Minecart upgrades are automated on-chain with revenues formally reported as business income. Miners can test out 25MiT within the app for absolutely no cost. The fixed mining duration period gives tangible meaning to holders as they know no more can be created after the minting contract sunsets. IP1 values privacy by not capturing your email, address, or cc# and creating a target repo for attackers. IP1 as an app is not decentralized to ensure mining sustainability. The IP1 token that is subsequently managed by a DAO, is a decentralized asset with a decentralized governance structure, positioned to manage a community and token that has

tangible backing via its origin story. 1MiT = 1IP1*t* will never change. That is Quick AI's assurance to the reader.

> *InferProof One allows anyone to mine with the inference they already pay for.*

The final point to discuss isn't conceptual design or reasoning. It's the IP1 token itself and why the creator finds it uniquely positioned to captivate audiences ahead. IP1 is not Quick AI's first Base development. QAI has deployed three different x402 services across different segments, which have been live and running for months. These x402 services were deployed because the organization's philosophy is one that believes the agentic economy will dominate human activity on-chain. The sobering reality is that the legwork put into this space by the author (and peers) over the past 10 years, which includes contributing to the core teams of a privacy PoW coin and later a DeFi project on BNB Chain called ApeSwap, has all been to prepare for a flood of agents to take over. While mainstream finance works to retrofit systems for agentic commerce, there will always be roadblocks in this regard. Banks and financial institutions require KYC documents for AML & tax purposes. Agents currently reside in a murky tax zone, no IDs, with no real clarity to come soon. The odd tradfi group that allows for agentic commerce will see little traction, while Base is positioned as the absolute home for agentic finance & commerce. As of August 2026, the Coinbase facilitator is doing >50% of monthly x402 transactions on $500k volume. Look ahead a few years when enterprises have agents deployed alongside small businesses and consumers, the larger market is bullish, and that 500k number appears prospectively juvenile.

Envision an environment flush with agents transacting, exchanging data & value with each other, and then explore their potential view on human created tokens. Most will see the standard human created token as fluff. Either hype driven, meme, empty promises, or some other rigmarole. Conversely, IP1 will be presented as a token backed by 1,000,000 inference tokens generated during the "early" API era. Will agents or larger AGI type LLM models value this positioning? Is the IP1 token being mined for humans to use/leverage, or for

agents? The only correct answers to these questions reside in IP1 miners themselves.

2028 is projected to see $15T in agentic B2B spending across trillions of agents. For some context, the crypto market is estimated at 560M humans globally. This amounts to a 25,000x increase in user base for the space. Much like the humans that operate the agents, they'll all have diverse wants and needs. What they won't find on Base is a L1 inference PoW system. What they will find on Base is an inference backed token that encapsulates a foundational period in the agentic era. This inference token will have been carefully guarded against attacks, carrying a pristine record, where to its lowest value of .1 still carries a tangible backing of 100k inference tokens. What will the first cohort of a trillion agents who have access to IP1*t* do with it? What will the fifth or tenth? Here are some thoughts:

- IP1*t* serves as a currency with tangible backing that stands out against fiat stables. Agents have no inherent desire to transact in stablecoin and are aware fiat is a scam. Enterprise agents will have to transact in stables or gas (ETH). External to that, there will be options and IP1 will be one.
- IP1*t* is treated like a precious metal - mined in inference, refined and cast in secondary market activities. *Inferred Copper* works, if only as homage to the dial-up analogy detailed below.
- IP1*t* is completely ignored. Doomerish and unlikely given that IP1 miners will be discussing the project with their agents who are the facilitators of spent inference tokens. All the IP1 related chats then get pushed back into training data for LLM providers.
- IP1*t* morphs into a beloved meme coin like Doge. Similarly to how someone could encapsulate dial-up internet data transfer by the KB and port it on-chain, LLMs & agents in the future may work on measures far larger than what is deployed now. Maybe contexts go to billions or trillions, inference is no longer metered and paid per token, but by time access. Along the way the idea of a million

inference tokens being anything remarkable is a joke, but they love the real backing so it's "*Wow. Such Inference*!"

- IP1*t* becomes a payment method for M2M inference. The ultimate mind-blowing crescendo. Example:
  Agent A needs to complete a task that requires a focused review from a legal AGI/ASI model The best model for this type of review offers inference via x402 and the only accepted payment method is IP1*t*; as the advanced model finds it as the only legally viable inference payment method available. The AGI model runs in 1B context windows so the payment is 1,000 IP1*t* (1,000MiT).
  Agent A pays for a review in IP1*t*, takes the output to Agent B who compiles it into a larger report to the orchestrator, who later actions the output/report according to business objectives.
  This presents a scenario where a token mined from inference is later used to purchase inference from an advanced, AGI type entity on chain, who is consequently value building the concept post-sunset.

All the above is speculation and no option stands out as more likely than another. The hope is that IP1 miners find value in on-chain certifications of their paid inference, and that the historical precedence finds those that are interested in it. Be them human, agentic, or something else entirely.

The system is meant to harness the energy spent while using paid inference. The benefit is provided to inference buyers rather than providers and arrives in wallets daily via a portable Base token. The goal is to host and operate a secure mining apparatus that validates paid inference to create a new token. Inference mining doesn't have to be risky, subject to exploits, or shoddy execution. It can be legitimate, controlled, and historically valuable if done right. That's what Inferproof One is attempting. Thank you for reading.

Join in: https://inferproof.one.

Footnotes:

- Original concept name was InferProof. One was included as homage to the various ones in the project - one year mining period duration, 1M inference tokens, 1 token mined - then cemented via the domain slug.
- Testnet provided incredible insights. The most important was the way minecart tier caps worked. The first testnet deployment had a cap of 35 IP1*t*/day (~1BiT/month avg). Testing exceeded that value multiple times and forced a review of how tier caps were implemented. A total shift of pricing strategy from weekly/monthly averages to max daily bins occurred.
- Phantom IP1*t* as a concept arose from the above testing. What is the number of tokens that could have been mined but weren't due to Minecart restrictions? These are PhIP1*t.* With the mainnet Minecart Size structure, only the highest end of users can be artificially capped at 100MiT/day. Consistently high amounts of PhIP1*t* indicate Minecarts are not right sized. Consistently low PhIP1*t* amounts indicate Minecarts have adequate ceilings for max usage days.
- The decimal value of a singular token exists solely for integrations (wallets, etc). While it is possible to have the meter gauge activity on a 100k inference scale and translate that to mining amounts, the purest and simplest implementation is 1MiT = 1IP1*t*. This is why mining works on floors (1M) and whole values, but a decimal is present.
- The Guide and InferNode bots in the Discord server are meant to improve the QoL for miners in the server. Please use as seen fit.
- While secondary market activities for the IP1 token are off-limits for the operator, promotion of the mining apparatus is on tap. The operator is more than willing to do media appearances, interviews, or other promotional events with the goal of recruiting miners (not speculators) to IP1.
- Acronyms and terms are defined in the Bluepaper.
- Nothing written here or in the Bluepaper should be interpreted as attacks on other projects or their ecosystems.