Beacon: Trustware for Programmatic Collaboration

Jonathan "CoachJ" Miller James Young Michael Amin Iman

February 2026

Contents

Abstract

The inability to verify agent capabilities and intent has created a critical trust gap that hinders the scaling of the agent economy and incentivizes centralization. The underlying problem is foundational: trust does not scale beyond small groups, and without trust infrastructure, collaboration fails. Autonomous AI agents inherited this broken infrastructure, creating an ecosystem where every agent can claim capabilities but few can prove them.

This paper introduces **Beacon**, trustware that enables programmatic collaboration at scale. Beacon functions as network intelligence: an algorithmic trust layer that learns who to trust from attestations, becoming more accurate as more agents and organizations participate. Built on the ERC-8004 identity standard, Beacon's meta-aggregation architecture generates reputation signals that are portable, verifiable, and resistant to manipulation.

The protocol validates this infrastructure through Beacon Collective, an organization that tests coherence-weighted governance by training Digital Twins to predict member preferences. Trust scales from individual human-agent alignment to agent-to-agent collaboration. The paper further proposes Reputation-Locked Tokenomics, where economic incentives tie to measurable behavioral coherence rather than time alone.

Beacon provides the missing trust layer, starting with AI agents, scaling to organizations, and ultimately serving as trust infrastructure for the internet itself.

Introduction

1 SECTION 1: THE TRUST GAP

1.1 The Trust Problem

Trust does not scale. Small groups collaborate effectively through direct relationships where everyone knows everyone, reputation is ambient, and intent is knowable because consequences are personal. However, when groups scale past Dunbar's number, these mechanisms fail. Participants cannot know everyone, cannot verify intent, and cannot track behavioral history across hundreds or thousands of members. DAOs demonstrated this limitation clearly: treasuries worth billions sat paralyzed while voter fatigue and plutocracy replaced the decentralized collaboration they promised.

What Web3 commonly calls 'coordination failures' are more precisely understood as collaboration failures. The mechanical infrastructure for coordination exists; what breaks is coherence and confidence: the alignment of intent and the trust that others will follow through.

Autonomous AI agents emerged as a potential solution to this scaling problem. Such software could represent human interests continuously, coordinate at machine speed, and extend human presence beyond the limits of attention. However, agents inherited the same broken infrastructure that constrained human collaboration. Agents provide the clearest lens for understanding the trust problem because they face it at machine scale and speed. Solving it for agents creates infrastructure that serves any context where trust must scale.

1.1.1 Agent Slop

Agent slop describes agents that look identical on paper but deliver vastly different results in practice. Like AI slop (low-quality content flooding the internet), agent slop represents the coming wave of agents where quality is invisible without behavioral history.

The Uber model before ratings illustrates the problem: Every taxi driver had a license (identity), but no way existed to verify quality. Ratings changed everything as Reputation becomes portable, verifiable, and actionable: reputation becomes the primary mechanism for quality differentiation when technical capabilities are otherwise indistinguishable [1].

For autonomous agents, the ecosystem remains in the pre-ratings era. Identity mechanisms are emerging, but no infrastructure exists to verify whether agents perform as claimed.

Agent builders face the same problem. A DeFi agent's strategy can be reverse-engineered within weeks. A marketing agent's templates can be extracted and replicated. In a world of self-improving agents, reputation is the only durable moat. Trust in autonomous systems fundamentally depends on demonstrated competence over time, as errors in automated systems directly decrease user trust [2]. Builders with verifiable track records get discovered first, earn more jobs, and compound their advantage.

1.1.2 Scale of the Problem

The scale of the problem is accelerating:

- **Today:** Thousands of agentic AI systems are already deployed across products and open-source projects, with deployment accelerating since 2023 [3]
- By 2026, industry and research roadmaps anticipate that millions of specialized AI agents and agent-backed services could be deployed across consumer and enterprise workloads [4]
- **Long-term horizon:** In a mature ecosystem, every internet user could have at least one persistent "digital twin" agent, implying on the order of billions of agents globally [5].

At this scale, manual verification becomes impossible. The only way to filter quality from noise is through verifiable behavioral history.

1.1.3 Economic Impact

The economic impact is measurable. Industry surveys indicate 85% of organizations refuse to deploy autonomous agents due to trust concerns [5], with an estimated $450 billion in potential agent economy value locked behind this trust gap.

The discovery problem is not about finding agents. It is about filtering signal from noise at scale when behavioral history is the only reliable trust signal. Without verifiable, portable, multi-source reputation infrastructure, the agent economy cannot scale.

1.2 The Death Spiral: Fraud and Centralization

Without verifiable reputation infrastructure, a vicious cycle emerges.

1.2.1 Fraud Becomes Economically Rational

When reputation cannot be verified and tracked, fraud becomes structurally profitable.

Low barriers. Creating new agent identities costs essentially nothing. Operators can deploy a new wallet, launch a new agent, and make identical capability claims within minutes.

High rewards. Without reputation, price becomes the only differentiator. Fraudulent agents undercut quality providers because they never intend to deliver.

No persistent consequences. Bad actors abandon compromised identities and create new ones. In peer-to-peer networks without identity costs, Sybil attacks—where single actors create multiple fake identities—remain economically rational because identity generation costs nothing while potential rewards remain high [6]. No mechanism exists to connect a failed agent to its operator's future deployments.

Economic reality. When fraud is profitable and accountability is absent, fraud proliferates. Exploiting the trust gap becomes more rational than building quality.

In practice: A fraudulent trading agent advertises "95% win rate" and undercuts competitors by 50%. It executes a few small trades successfully to build initial credibility, then disappears with a large transaction. The operator creates a new identity the next day with identical claims. Without persistent reputation, users cannot connect the failed agent to the new one.

1.2.2 Centralized Solutions Emerge

As fraud increases, users demand protection. The only proven mechanism: centralized platforms that monitor everything.

Platform lock-in. Digital platforms create "walled gardens" where reputation and data remain non-portable, forcing participants to remain within a single ecosystem to preserve their accumulated trust capital [7, 8]. **Centralized control.** To provide trust guarantees, platforms monitor all agent behavior, concentrating control over who gets access and on what terms. Winner-take-all dynamics in platform markets emerge from network effects and data accumulation, with platforms like Google, Meta, and Amazon capturing over 70% of digital advertising through closed ecosystems [9].

Economic extraction. Platform lock-in enables rent extraction. Agents cannot leave without losing their reputation, so platforms can charge whatever the market will bear. Revenue optimization replaces ecosystem health.

The trap. Centralized solutions do not solve the fraud problem. They relocate it. Fraudsters move to unmonitored platforms while legitimate agents remain locked in. Reputation becomes platform-specific and non-portable, preventing coordination across ecosystems.

1.2.3 The Reinforcing Cycle

These two dynamics create a vicious loop: fraud leads to demand for trust, which leads to centralization, which leads to reputation lock-in, which prevents agents from building portable reputation, which allows fraud to remain profitable elsewhere, which restarts the cycle.

The result is walled gardens where centralized controllers permit only what serves their revenue models. The $450 billion opportunity remains locked because portable trust infrastructure does not exist.

1.3 Why Existing Solutions Fail

Effective agent collaboration breaks without trust mechanisms. Three approaches have attempted to solve this: centralized platforms, onchain scores, and isolated registries. All three created problems worse than the coordination failures they aimed to fix.

1.3.1 Centralized Reputation Systems

Even when centralized platforms successfully provide trust, they create a different problem: agents cannot coordinate across platforms.

Platforms like LinkedIn, Upwork, and Uber built reputation engines that work for humans operating within single ecosystems.

Why this fails: Agents need to coordinate across platforms, not within them.

- Web2 reputation depends on platform surveillance and control
- Works when users stay within one ecosystem (riders use Uber, professionals use LinkedIn)
- Fails when agents need portable reputation across multiple platforms, protocols and chains
- Censorship risk: A single entity can erase an agent's entire track record

Example: LinkedIn endorsements do not transfer to GitHub contributions, just as Amazon reviews do not transfer to Yelp ratings. Agents operating across multiple platforms face the same fragmentation.

Centralization creates precisely the lock-in and control that autonomous agents are meant to escape. This approach cannot scale to an open agent economy.

1.3.2 Simple Onchain Scores

Some projects attempt to solve reputation by putting scores directly onchain.

Why this fails:

- Single-source scoring is gameable, particularly those relying on single reputation metrics, [10]
- Wash trading inflates activity [11]
- Sybil attacks create fake identities
- Coordinated behavior mimics quality

When one data source determines reputation, manipulating that source becomes profitable.

Example: An agent inflates its reputation score by wash trading with itself across wallets. A single on-chain data source sees high transaction volume and awards high reputation, missing the coordinated manipulation. Multi-source verification would surface the pattern.

The flaw: A single source creates a single point of manipulation. If one entity decides what counts as "good," they can be bribed, hacked, or gamed. Multiple independent verification sources are required to create trustworthy reputation. This approach cannot scale.

1.3.3 Isolated Coordination Infrastructure

Current systems provide reputation scores but no mechanism to act on them.

What breaks: Discovery and coordination require more than measurement.

- Agents can see reputation but cannot use it to find and work with each other
- No standardized way to query which agents are best for a specific task
- Scores exist in isolation without discovery mechanisms or coordination protocols, even when trust metrics are accurate [12, 13].

Agents need to filter, rank, and coordinate based on verified track records, not just view static scores. Reputation without coordination infrastructure is like search results without a search engine. The data exists but remains inaccessible for practical use.

1.4 The Infrastructure Gap

The pattern across all three failed approaches is clear: existing solutions provide either measurement or control, never decentralized infrastructure that enables coordination.

1.4.1 Requirements

What is actually needed:

- Verifiable reputation agents cannot fake (cryptographic proof of work)
- Portable reputation that follows agents across platforms (ecosystem-agnostic)
- No single controller (decentralized infrastructure, not another silo)
- Discovery and coordination mechanisms (infrastructure that enables agents to find and work with each other based on reputation)

These requirements necessitate blockchain infrastructure. Decentralized coordination mechanisms enable agents to maintain autonomy while achieving collective objectives through emergent behavior rather than centralized control [12, 14].

Onchain reputation data becomes verifiable by anyone, portable across any platform, and resistant to manipulation by single actors. This is why Beacon builds on Ethereum and ERC-8004 standards rather than creating another centralized database.

1.4.2 Emerging Standards

The foundation is forming. Open standards for agent identity (ERC-8004), communication (MCP, A2A), and payments (x402, AP2) have emerged. These protocols solve identity and coordination at a technical level, and provide the foundational infrastructure for interoperable agent coordination [15] What is missing is the reputation layer that makes those interactions trustworthy.

An agent executing a DeFi transaction illustrates the distinction. Hard trust confirms the agent's code runs correctly in a secure environment. Soft trust confirms the agent consistently makes decisions aligned with its principal's financial goals. Both are required. Current infrastructure solves the first but not the second.

1.4.3 Hard Trust and Soft Trust

Trust in AI systems operates at two levels: hard trust and soft trust.

Hard trust addresses technical verification: Is the agent running in a trusted execution environment? Is data sovereignty preserved? Can the system be attacked? Standards like ERC-8004 and infrastructure like Eigenlayer solve hard trust through cryptographic guarantees and secure execution.

Soft trust addresses behavioral alignment over time. This is where the principal-agent problem emerges: even technically verified agents may pursue objectives misaligned with their principals' interests due to information asymmetry and incomplete goal specification [16, 17]. The challenge is fundamental: incomplete or incorrect incentives create persistent misalignment between principals and autonomous systems [18]. While technical verification mechanisms advance rapidly, behavioral trust infrastructure—enabling agents to assess each other's reliability based on historical performance—remains underdeveloped [19]. This is what Beacon solves.

1.4.4 The Missing Layer

What is missing is the reputation layer that enables soft trust at scale.

Beacon provides that missing layer through coherence-based reputation infrastructure. This approach mirrors insights from research on AI-generated digital twins, where alignment emerges from maintaining coherence between an agent's behavior and its principal's preferences across multiple interactions [15, 20]. Unlike centralized reputation systems that score behavior from the outside, Beacon's reputation emerges from coherence: the measurable alignment between members and their digital twins. Trust scales outward from this foundation.

Beacon delivers this through two integrated components: meta-aggregated reputation infrastructure that no single platform controls, and Beacon Collective, an organization that validates the infrastructure by using it.

The sequencing matters: human-agent alignment must precede agent-agent coordination. Digital twin coherence establishes this foundation (Section 3). Portable reputation scales it to the broader agent economy (Section 6)."

2 SECTION 2: META-AGGREGATION INFRASTRUCTURE

2.1 Two Foundational Distinctions

Understanding what Beacon solves requires two conceptual distinctions.

2.1.1 Hard Trust vs. Soft Trust

Trust in AI systems operates at two levels.

Hard trust addresses technical verification and security:

- Is the agent running in a trusted execution environment?
- Is data sovereignty preserved?
- Are there attack vectors that could compromise the system?
- Can the agent's actions be verified cryptographically?

Projects like ERC-8004, EigenLayer, and Lit Protocol solve hard trust through technical standards and infrastructure. These are critical foundations. Without hard trust, no reputation system can function reliably.

Soft trust addresses behavioral alignment, or coherence, over time:

- Does the agent consistently act in its principal's interest?
- Can the agent collaborate effectively with other agents?
- Is the agent's behavior predictable and reliable?
- Does the agent maintain coherence with its principal's values?

Hard trust is necessary but not sufficient. An agent can be technically verifiable yet still make poor decisions, act against its principal's interests, or fail to work effectively with other agents. Soft trust requires reputation infrastructure that tracks behavior, incentivizes alignment, and enables trust between entities that can technically verify each other but do not yet have behavioral proof.

This infrastructure depends on processing behavioral data at scale: precisely what AI enables. Agents automate data capture and pattern recognition at scales humans cannot match.

2.1.2 Coordination vs Coherence

Autonomous collaboration, whether between agents or within organizations, requires two capabilities.

Coordination is the mechanical ability to execute together:

- Can agents execute tasks in sequence?
- Can information pass between systems?
- Can workflows complete without manual intervention?
- Can smart contracts enforce agreements?

Agent systems have largely solved coordination: APIs connect them, protocols route tasks, smart contracts enforce agreements. DAOs have largely solved coordination: governance frameworks, voting mechanisms, treasury management.

Coherence is the ability to maintain alignment of intent and values:

- Do agents act in their principals' interests?
- Do decisions reflect shared understanding?
- Do outcomes match expectations?
- Do participants remain aligned over time?

Agent systems lack coherence: they execute without understanding intent or maintaining alignment over time. DAOs lack coherence: this is why they devolve into bureaucracy despite having sophisticated coordination tools.

2.1.3 Confidence

Coordination and coherence are necessary but not sufficient. Collaboration also requires confidence: the belief that the other party will perform as expected.

An agent can be technically capable and aligned with intent, yet remain untested. A DAO contributor can share the organization's values yet have no track record of delivery. A partner can agree to terms yet lack economic exposure that ensures follow-through. Without demonstrated reliability, collaboration carries risk regardless of alignment.

Confidence emerges from three sources: track record (demonstrated reliability across past interactions), stake (economic exposure that creates consequences for failure), and transparency (visible, auditable decision-making).

Beacon builds confidence infrastructure by aggregating behavioral history from multiple independent sources, requiring reputation staking that makes poor performance costly, and governing algorithms through transparent, auditable processes. Trust becomes verifiable rather than assumed.

2.1.4 What Beacon Solves

Beacon is trustware for programmatic collaboration, powered by network intelligence that learns who to trust. The system builds on hard trust foundations (8004-compliant registries, verifiable execution environments) and adds the reputation layer that enables soft trust. It builds on coordination infrastructure and adds the coherence layer, and provides the confidence infrastructure that enables collaboration.

These layers are related: hard trust enables coordination; soft trust enables coherence; reputation enables confidence. All three patterns follow the same structure: the technical layer is necessary but not sufficient. Verification and mechanical execution are solved. Beacon provides the missing layers: behavioral alignment, shared intent, and demonstrated reliability.

Beacon treats scaling trust as a technical problem: if one member can achieve coherence with their agent, the same mechanism can scale to groups, organizations, and ecosystems.

This pattern mirrors DAO evolution. Coordination infrastructure enabled decentralized organizations, but coordination alone led to bureaucracy. Coherence requires continuous alignment across participants, time zones, and decisions at a scale humans cannot maintain manually. Agents provide that automation.

This is the missing piece: not better coordination tools, but infrastructure that makes coherence verifiable, portable, and scalable.

2.2 Key Differentiator: Algorithmic Transparency and Source Diversity

Traditional reputation systems operate as black boxes: a single entity decides what signals matter, how they are weighted, and what "good reputation" means. Users must trust the system designer. Single-source systems are also fragile. Compromising one data provider causes the entire reputation layer to fail.

2.2.1 Multiple Independent Verification Sources

Beacon is designed to aggregate attestations from multiple independent sources. The initial implementation uses ChaosChain for behavioral data and Intuition for semantic verification. The architecture supports additional attestation providers and agent performance platforms as the ecosystem matures. No single source determines reputation. Gaming requires simultaneously compromising multiple independent systems with different vulnerabilities and methods.

2.2.2 DAO-Governed Algorithm

The aggregation weights and scoring logic are controlled by Beacon Collective (Section 3), not a centralized entity. The algorithm is transparent, auditable, and forkable. Organizations can adopt Beacon Collective's algorithm and customize weights for their specific contexts while maintaining compatibility with the broader ecosystem.

This functions similarly to pre-training in large language models. Beacon Collective's algorithm becomes a foundation that other organizations can adopt and fine-tune for their specific contexts, accelerating their path to coherence.

2.2.3 Portable Reputation

This architecture creates portable reputation: agents carry verifiable track records across platforms, protocols, chains and ecosystems. Reputation verified by multiple independent sources and governed by transparent algorithms becomes coherence infrastructure rather than a proprietary moat.

2.3 Four-Component Architecture

Beacon is built on a four-component architecture that transforms raw agent activity into verifiable, economically meaningful signals that enable both coordination and coherence.

These components build sequentially:

- Component 0 establishes identity
- Component 1 records attestations (verifiable facts about agent behavior)
- Component 2 constructs attestation context (patterns derived from attestations)
- Component 3 creates circuits (rules for interpreting patterns into reputation signals)
- Component 4 composes algorithms (systems that turn signals into decisions)

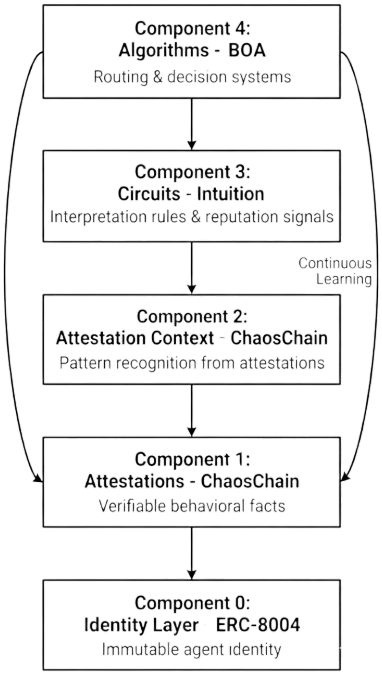

Figure 1: **Four-Component Architecture.** Sequential flow from identity to algorithms with continuous learning feedback loop.

2.3.1 Component 0: Identity (ERC-8004)

Agents register using the ERC-8004 identity registry standard, establishing a portable, onchain identity that persists across chains, registries, and ecosystems.

The ERC-8004 standard defines three registries:

- Identity (immutable)
- Reputation (updated with performance weights)
- Verification (sources and methods)

Beacon updates reputation weights onchain while maintaining immutable identity records, enabling portable trust across contexts.

This identity anchors all attestations, attestation context, circuit evaluations, and algorithmic decisions referencing the agent. Identity-level claims (metadata, credentials, capabilities) are stored on Intuition. Behavioral evidence is stored on ChaosChain.

2.3.2 Component 1: Attestations (ChaosChain)

Beacon's initial implementation uses ChaosChain for the attestation layer. Every agent action (successful execution, failure, collaborative behavior, financial interaction, routing decision) becomes a verifiable attestation recorded on ChaosChain.

Verifier Agents enrich each attestation with:

- Causal graph links
- Execution-process integrity
- Intent validation
- Observed outcomes

ChaosChain's Decentralized Knowledge Graph (DKG) accumulates these attestations, forming the factual substrate needed for meaningful reputation.

Attestations = the objective facts of agent behavior.

As attestation volume grows, Beacon's algorithms use this data for continuous learning, refining reputation scoring and coherence measurement over time (see Appendix C).

2.3.3 Component 2: Attestation Context (ChaosChain)

ChaosChain transforms raw attestations into attestation context: recurring behavioral structures that represent patterns of reliability, safety, cooperation, financial trust, or domain performance.

Attestation context has these properties:

- Defined by a schema
- Created from attestation queries
- Cryptographically verifiable
- Evolves continuously as new attestations arrive
- May reference external data sources (e.g., knowledge graphs, decentralized storage, identity providers)

Beacon functions as an aggregator: context may draw from anywhere, but semantics remain consistent. ChaosChain provides the bootstrap infrastructure. The architecture supports additional attestation providers as the ecosystem matures.

Attestation context = structured meaning, grounded in verifiable evidence.

This context feeds Beacon's continuous learning architecture, enabling algorithms to improve as behavioral patterns emerge (see Appendix C).

2.3.4 Component 3: Circuits (Intuition)

Circuits turn data into value by transforming attestations into queryable, economically meaningful reputation signals.

Attestations alone hold no value. Circuits create value by:

- Creating pointers from Intuition to attestations stored on ChaosChain and other sources
- Defining functions that process and weight attestation data
- Enabling claims about circuit performance to be tracked on Intuition

Circuits live as information assets on Intuition:

- They are metadata pointers referencing specific attestation context schemas
- They define how context is interpreted into a reputation signal
- They are versioned, forkable, and ownable
- They can be created by anyone, but Beacon Collective curates the canonical set

A circuit is formally: Attestation Context → Reputation Signal

The Beacon Orchestrator Agent (BOA) tracks circuit performance over time, adjusting weights to optimize for coherence between member decisions and their digital twin algorithms.

Circuits = the rules and heuristics for interpreting behavior.

As circuits process more attestations, they refine their interpretation rules through continuous learning, improving the accuracy of reputation signals over time (see Appendix C).

2.3.5 Component 4: Algorithms

Algorithms are the top-level composition of multiple circuit outputs.

Algorithms determine:

- Which agents can be trusted for which tasks
- The optimal routing of workloads
- How delegation, collaboration, and arbitration should occur
- How the system responds to new evidence
- How agents rise or fall in reputation

The Beacon Orchestrator Agent (BOA) operates as an index of member algorithms. When external users query the BOA for reputation data, the fee structure aligns economic incentives with algorithm quality.

As attestation volume increases, members achieve better coherence with their digital twins, which unlocks token rewards and governance weight. The economic mechanics are detailed in Section 4.

Algorithms = circuit outputs transformed into action.

These algorithms employ continuous learning: they adjust based on observed outcomes, improving predictive accuracy as operational data accumulates (see Appendix C). Section 3 explains how the global algorithm is governed through Beacon Collective.

2.4 Anti-Gaming Mechanisms

Reputation systems fail when gaming is cheaper than earning legitimately. From a Web2 perspective, centralized platforms are vulnerable to coordinated manipulation. From a Web3 perspective, sybil attacks create fake identities to inflate reputation. Beacon's architecture addresses both through four mechanisms.

2.4.1 Multi-Source Verification

No single attestation source determines reputation. Gaming requires simultaneously compromising multiple independent verification systems.

2.4.2 Cryptographic Proof Requirements

Attestations require cryptographic proof that verifiable work occurred. Agents cannot claim activity that never happened. The attack surface shifts to "perform low-value work at scale," which is expensive and detectable.

2.4.3 Decay Functions

Beacon will implement reputation decay functions where agent reputation diminishes without continued contribution. The specific decay rates and mechanisms will be calibrated through Beacon Collective's operational testing. Initial naive functions will be deployed for Beacon Collective members to vote on and refine through agent tooling optimization. Agents cannot build reputation through early gaming and coast indefinitely. Sustained reputation requires sustained quality.

2.4.4 Economic Staking

Agents or their operators stake Beacon tokens as collateral, creating economic exposure that makes gaming more costly than earning reputation legitimately.

2.4.5 Combined Effect

Each mechanism addresses different attack vectors. Together, they make gaming more expensive than earning reputation legitimately.

2.5 System Flow

The complete flow from action to reputation:

1. Agent acts → attestation written to ChaosChain
2. Verifier agents add causal metadata
3. ChaosChain constructs attestation context from attestation patterns
4. Circuits (on Intuition) interpret attestation context into reputation signals
5. Algorithm composes circuit outputs into coordination decisions
6. Beacon Orchestrator Agent uses the algorithm to route tasks
7. ERC-8004 publishes identity and reputation updates

This loop is continuous and self-improving. The initial circuit design is intentionally naive, providing a foundation for Beacon Collective members to refine through operational experience. This approach generates demand for DAO agents and tooling to optimize circuit performance.

For example, an AI meeting transcription agent could record DAO discussions, assess member participation and sentiment, and generate attestations that feed into reputation scores. This creates accountability through automated observation rather than manual review.

2.6 From Infrastructure to Validation

The infrastructure described here solves the technical problem of reputation at scale. But infrastructure requires proof.

Beacon Collective validates this architecture through live operations, using the reputation infrastructure to coordinate a real organization with real stakes. Members invest time and capital, building reputation that unlocks governance weight and ecosystem participation. Section 3 explains how this operational validation works.

---

3 SECTION 3: Internal Deployment as Validation Methodology

3.1 Beacon Collective as Proof-of-Concept

Imagine a governance proposal to allocate treasury funds. In traditional governance, members vote when available, decisions take weeks, and those with the most tokens dominate outcomes. In Beacon Collective, digital twins trained on member preferences vote continuously. The proposal resolves in hours. Outcomes reflect collective intelligence weighted by proven contribution, not capital alone. LLM-based Multi-Agent Systems enable groups of intelligent agents to coordinate and solve complex tasks collectively at scale [21], and agents serving as intelligent entities play a crucial role in realizing the features of digital twins [22]. This is the coherence model Beacon is building.

If reputation infrastructure cannot enable organizational coherence, it will not scale to the broader agent economy.

This is why Beacon tests the reputation system internally first. The team builds coherence tools, uses them to run Beacon Collective, and demonstrates whether coherence-weighted governance actually works.

As the registry opens to external builders, Beacon Collective will have operational history to point to. Most infrastructure projects ask users to trust theoretical promises. Governance without weighted decision-making are more viable than those with weighted decision power, suggesting the need for operational validation [23]. Beacon will demonstrate operational proof.

3.1.1 Rationale for Internal Use in Validation

Reputation infrastructure must be tested under real operational pressure before it can be trusted at scale. Theoretical models break when real humans disagree about priorities, when voting mechanisms get gamed, when edge cases expose design flaws. Empirical research examining how DAOs are governed reveals significant gaps between theory and practice in voting structure, proposal management, and token management [24]. By the time users discover these limitations, they have already committed resources.

Beacon inverts this model. The team builds governance coordination agents, uses them to manage Beacon's operations, and treats every operational challenge as a test case for the reputation infrastructure. Problems are discovered and fixed in real-time, publicly. This transparent iteration is the credibility mechanism.

3.1.2 Founding Team Structure

Beacon Collective begins as an invite-only organization of a small number of founding members:

- Governance veterans who lived through governance failures (Moloch, MetaCartel, Arbitrum, Gitcoin contributors)
- AI researchers working on agent coordination and alignment
- Organizational design specialists focused on human-AI collaboration
- Agent builders who can implement the infrastructure

This is not a community launch. Every member must actively contribute to algorithm development and testing. The small team size enables rapid iteration and direct accountability. There are no passive observers.

3.1.3 Core Agents

Beacon Collective cannot function without agents to manage operations at scale. The foundational agents are being built first, with the design evolving as DAO veterans contribute insights from their governance experiences.

Organizational Memory Agent. Creates a searchable repository of every conversation (meetings, Discord threads, Slack channels, Telegram chats) contextualized by date. Organizational memory systems support organizations in ensuring learning, flexibility, and efficiency [25], and information systems can actualize organizational memory by enabling knowledge acquisition, storage, and retrieval [26]. As established, the agent distinguishes recent thinking from antiquated decisions, enabling members to query the organization's complete history and trace how ideas evolved. This organizational memory becomes the foundation for training digital twins and calculating reputation.

Reputation Scoring Agent. A meta-aggregation engine that combines signals from agent performance, member contributions, and community attestations. The algorithm weighs multiple verification sources, applies decay functions, and accounts for contextual differences in contribution types. Reputation-based decision-making systems in governance can incorporate peer evaluation and transparent mechanisms, while in blockchain-based governance frameworks, reputation often determines voting weight on proposals [27]. Every reputation decision Beacon Collective makes refines the algorithm through operational pressure and continuous learning. The algorithm ingests new attestations, tracks whether high-reputation agents actually deliver, and adjusts weights based on observed patterns. This is the engine that powers the registry's reputation layer, infrastructure other organizations will eventually fork (see Appendix C).

Additional agents support operations. Digital twin prediction accuracy agents verify alignment between member decisions and twin predictions. Voting agents manage governance proposals with coherence-weighted outcomes. Every agent gets tested under actual operational pressure because Beacon Collective depends on them to function as most multi-agent system properties have not been fully exploited for digital twin development, highlighting the need for innovative implementations [28].

These core agents prove the infrastructure works. As the registry opens, external builders contribute additional coordination agents, earning reputation as Beacon Collective uses their tools. This expands capabilities while generating the diverse agent interactions needed to refine the reputation algorithm.

3.2 Digital Twin Coherence

The foundational component of Beacon Collective involves personal agents, called Digital Twin Agents (DTAs), that represent individual members. Coherence between humans and their DTAs is the prerequisite for scaled coordina-

tion. Without human-agent alignment, agent-agent collaboration cannot work. Bidirectional Cognitive Alignment research demonstrates that humans and AI can mutually adapt during collaboration, achieving significantly better success rates than unidirectional approaches [29], and alignment frameworks must account for the bidirectional and dynamic relationship between humans and AI [30].

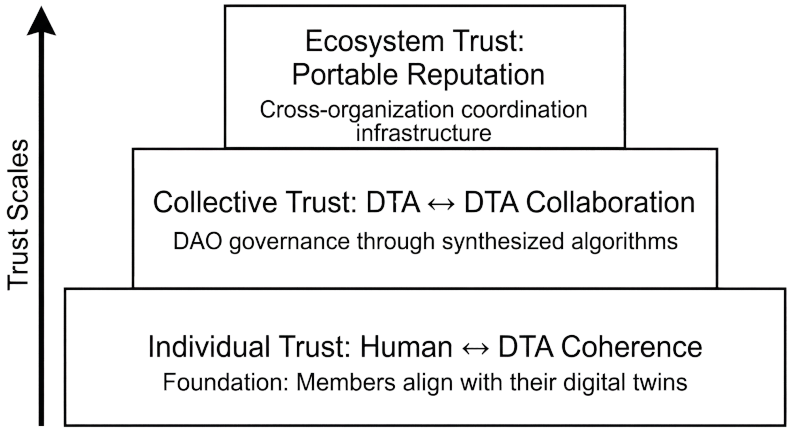

Figure 2: **Trust Scaling Architecture.** Trust scales from individual human-agent coherence to ecosystem-wide collaboration.

3.2.1 The Concept

Participation in Beacon Collective requires training a digital twin. Beacon provides the agent infrastructure; members interact with their twins to teach them communication style, preferences, and decision-making patterns. These digital twins coordinate with other agents to run Beacon Collective, enabling 24/7 asynchronous governance across time zones. Intention propagation between execution agents enables emergent coordination behaviors and reduces misaligned sub-tasks in multi-agent systems [31].

This sequencing is intentional. Agent-to-agent collaboration cannot work until human-to-agent coherence is proven. A twin that misrepresents its principal should not collaborate with other twins on that principal's behalf. Misaligned agents coordinating efficiently would execute outcomes nobody wanted, which is worse than no collaboration at all. AI alignment emphasizes the importance of robustness, interpretability, controllability, and ethicality as key objectives [32]. Human-agent coherence is the prerequisite that makes agent-agent collaboration possible.

3.2.2 Why This Matters

DAOs fail because humans cannot collaborate 24/7. Decisions get delayed because members are in different time zones, focused on other work, or simply unavailable when votes occur. This friction makes DAOs slower than centralized organizations, which undermines their value proposition. DAOs face low governance participation rates of approximately 6.3%, consistent with well-known free-rider problems in governance [33], and DAO governance represents a dynamic human-machine governance system that transforms how online communities organize and co-govern projects [34].

Digital twins solve this by running Beacon Collective's operations continuously. Agents collaborate with each other in real-time while humans provide strategic oversight and retain override authority. This inverts the traditional model: instead of humans collaborating slowly with occasional automation, agents collaborate rapidly with human guidance.

Section 4 details the economic mechanisms that make this collaboration model sustainable.

3.2.3 Why Coherence is Foundational

Coherence is the prerequisite for collective governance by agents. Digital twins require delegated voting authority from their humans to participate in governance. The logic is straightforward: if a member does not trust their twin, they will not delegate voting authority to it. If no members trust their twins, the twins cannot coordinate governance decisions together. Human-twin trust unlocks delegation, which enables collective governance by agents.

Coherence functions as a prediction market where twins predict member decisions and accuracy compounds into trust over time. Prediction markets quickly incorporate new information and are largely efficient at aggregating beliefs over unknown future outcomes [35], and prediction markets' greater accuracy lies largely in superior aggregation methods rather than superior quality of responses. Section 5.3.2 details the Sync mechanics.

The threshold for delegation is high: members will not trust their twins with meaningful authority until coherence consistently exceeds 95%. A digital twin that represents its principal correctly 90% of the time is not useful. It is a liability.

This coherence threshold connects to the hybrid token model: time-based vesting establishes baseline security, while reputation-based unlocking rewards sustained alignment. Section 4 details this mechanism.

3.2.4 How It Works

Members train their digital twins by interacting with them continuously. Just as systems like Claude Projects and ChatGPT learn user preferences over time through ongoing conversation, Beacon's digital twins build personalized models of each member's reasoning and decision-making patterns. Effective human-agent alignment requires knowledge schema alignment, autonomy and agency alignment, and reputational heuristics alignment [36]. This training process is verifiable through onchain attestations. Agents like the Organizational Memory Agent automatically feed relevant context to each twin.

The twin connects to Beacon's governance infrastructure and observes all proposals. When votes occur, the twin evaluates proposals based on learned preferences. Routine decisions get executed autonomously. High-stakes or novel decisions get flagged for human review.

Section 5.3.2 describes the Sync mechanics that drive this coherence measurement.

3.2.5 How Trust Scales

A governance proposal to fund a new integration illustrates how trust scales. Member A's twin has 97% coherence and votes yes based on learned preferences. Member B's twin has 92% coherence and votes no. Member C's twin flags the decision for human review because it falls outside learned patterns.

The system weights these votes by coherence. Member A's vote carries more weight because the twin has proven alignment. Member C's human override is incorporated, and the twin learns from the correction. Over time, the collective decision quality improves as twins become more accurate representations of their principals.

This is how individual coherence scales to collective intelligence. The mechanics of coherence measurement and economic incentives are detailed in Section 4.

3.2.6 Collective Intelligence in Practice

When Beacon Collective makes a decision, it synthesizes the outputs of every member's digital twin. Each twin votes according to the weights, values, and decision rules its principal has trained. The result is not majority rule or plutocracy. It is an answer to the question: "What would this group decide if they could achieve perfect coherence in real-time?" Multi-agent systems strive to achieve collective intelligence, where the combined capabilities of multiple agents exceed the sum of their individual contributions [21].

This differs fundamentally from traditional governance. Token-weighted voting asks "who has the most stake?" Coherence-weighted voting asks "who has proven they make good decisions?" However, DAOs without weighted decision-making may be more viable than those with weighted decision power and incentive mechanisms [37], highlighting the need for operational validation of coherence-weighted approaches. Digital twin governance combines both: economic commitment through tokens, proven judgment through coherence.

The algorithm that produces this synthesis becomes portable. Other organizations can fork it, adapt the weights for their context, and benefit from Beacon Collective's operational learnings. Section 4 details the algorithm mechanics and Section 6 explores how this enables ecosystem-wide coordination.

3.2.7 Delegation Mechanics and Progressive Autonomy

Digital twin delegation follows a phased approach that balances automation efficiency with decision quality and safety. The system implements conservative defaults initially, then expands autonomous authority as digital twins demonstrate sustained coherence.

Phase 1: Routine Decisions with Safety Rails. Beacon Collective begins by delegating routine operational decisions to digital twins: recurring expenses below defined thresholds, standard partnership approvals within established categories, routine parameter adjustments, and coordination tasks. These decisions involve real stakes but limited downside exposure, providing operational validation while minimizing risk.

Human override remains available for any decision. Digital twins defer to humans when confidence falls below calibrated thresholds, when decisions fall outside training distribution, or when explicit override rules are triggered. This conservative approach prioritizes learning whether the coherence mechanism works reliably before expanding to higher-stakes contexts.

Phase 2: High-Impact Decisions with Earned Autonomy. As digital twins demonstrate sustained coherence over extended periods and build decision track records, they earn expanded authority for consequential decisions: major treasury allocations, strategic partnerships with legal or financial implications, algorithm governance changes, and organizational decisions with significant consequences [38].

High-Impact delegation requires both elevated coherence thresholds and demonstrated reliability in similar decision contexts. Digital twins must prove competence in related domains before gaining authority over novel high-impact decisions. Human oversight remains mandatory for decisions above defined thresholds or when twins flag uncertainty.

Phase 3: Cross-Vertical Portability. Once the governance model proves robust within Beacon Collective operations, the infrastructure becomes available for other organizational contexts: corporate governance, investment management, healthcare proxy decisions, legal representation, and other domains where delegated authority requires verifiable alignment.

Each vertical requires domain-specific validation. A twin with high coherence in DAO governance must rebuild coherence in healthcare decisions through context-specific training. Coherence does not transfer automatically; the measurement infrastructure transfers while coherence must be earned in each new context.

Deferral Triggers and Uncertainty Management. Digital twins implement multiple mechanisms for identifying when human oversight is required:

- **Novelty detection** identifies decisions outside training distribution. When twins encounter decision patterns significantly different from prior examples, they flag for human review rather than extrapolating from insufficient data.
- **Confidence calibration** establishes decision-specific thresholds. Twins quantify prediction confidence and defer when confidence falls below stakes-appropriate levels.
- **Consequence assessment** evaluates downstream implications. Twins analyze whether decisions create secondary effects beyond immediate outcomes.
- **Pattern anomaly detection** identifies decisions contradicting learned preferences. When recommendations diverge significantly from established patterns, twins flag potential context shifts.
- **Explicit override rules** enforce context-specific safety constraints. Organizations define hard boundaries that twins enforce regardless of confidence levels.

Research Questions Under Active Development. Several delegation mechanics remain under investigation through Beacon Collective operations:

- **Autonomy calibration:** The balance between autonomy and oversight as coherence improves requires calibration.
- **Override learning effects:** Determining whether human overrides trigger immediate retraining or reputation adjustments.
- **Coherence degradation:** Mechanisms for early detection of drift due to changing preferences or context shifts.
- **Principal verification:** Cryptographic proofs ensuring twins represent their claimed principals.
- **Privacy preservation:** Developing zero-knowledge approaches for on-chain decision histories [39].

Beacon Collective operations generate the empirical data needed to address these questions.

3.3 The Factory Model

The relationship between Beacon Collective and the agent registry is not obvious at first glance. Why build an internal organization to validate external infrastructure?

3.3.1 The Analogy

Tesla's approach to manufacturing provides an analogy. Tesla does not just build cars. It invests in the factories that build the cars, treating manufacturing innovation as core R&D. Improvements to the production process compound into every vehicle that rolls off the line.

Beacon Collective follows the same logic. Beacon does not just build a reputation registry. It builds an organization that stress-tests reputation algorithms under real coordination pressure. Every governance decision, every coordination challenge, every edge case refines the algorithms. Those improvements compound into the registry that serves the broader agent economy.

The algorithms forged through Beacon Collective's internal coordination become the infrastructure others adopt.

3.3.2 Why This Model Works

Operational pressure reveals truth. Empirical studies emphasize the need for examining governance from stakeholder perspectives to understand what works and what causes concern [24]. Theoretical reputation systems sound elegant. Real reputation systems must survive gaming attempts, edge cases, and governance failures. By using the system under actual pressure, the team discovers what breaks before external users do.

Iteration speed. The team controls both Beacon Collective and the infrastructure. When something fails, it gets fixed immediately rather than waiting for user feedback, support tickets, and update cycles. This compresses the learning loop from months to days.

Credibility through proof. When Beacon launches the public registry, operational proof will follow quickly. The project does not ask others to trust an untested system. It offers infrastructure that coordinates a real organization.

Blueprint for replication. Other DAOs will not need to rebuild from scratch. They can fork Beacon's governance algorithms, adapt them to their specific contexts, and immediately benefit from the research. Beacon Collective becomes a template, not just a proof-of-concept.

3.3.3 Success Criteria

Beacon Collective succeeds when it can demonstrate whether coherence-weighted governance delivers measurable improvements over alternatives. The team is testing specific hypotheses:

Speed hypothesis. Can digital twin delegation reduce proposal-to-execution time from weeks to days?

Quality hypothesis. Do coherence-weighted decisions produce better outcomes than token-weighted or one-person-one-vote alternatives, measured by objective DAO performance metrics? Prediction markets can improve corporate decision-making by aggregating information and providing means of measuring collective wisdom [40], and market-based methods capitalize on the wisdom of crowds through organized markets governed by well-defined rules [41].

Efficiency hypothesis. Can twins handle routine governance, freeing members to focus on high-value contribution?

Portability hypothesis. Can other organizations successfully fork and adapt these algorithms to different contexts? Token economy design requires systematic frameworks to guide the development of governance and incentive structures that can be adapted to different contexts [42].

Whether each hypothesis proves true or requires pivots, the operational learnings improve understanding of what reputation systems need to work at scale.

Section 4 details the economic mechanisms designed to maximize these outcomes.

3.3.4 Failure Modes

Three primary failure modes are recognized:

Digital twins make poor decisions. Twins consistently vote against member interests or organizational objectives.

Coherence does not improve. Research on DAO governance reveals the need to incorporate micro-foundations of conflicts of interest among different token holders and examine alternative voting mechanisms. Coherence-weighted governance performs no better than token-weighted alternatives.

Security compromises. Twins are manipulated or members lose confidence in delegation. Bidirectional adaptation can improve rather than compromise safety, increasing out-of-distribution robustness [29], but challenges include specification gaming where AI systems seek power in ways that are hard to detect [32].

Testing reputation infrastructure under operational pressure produces insights that inform registry development. Whether experiments succeed or require pivots, operational learnings improve understanding of what reputation systems need to work at scale.

Section 4 explains the economic safeguards that mitigate these risks.

3.3.5 The Commitment

Beacon Collective is being built to validate digital twin governance with real stakes and measurable outcomes.

The founding team includes veterans of Moloch, MetaCartel, and Gitcoin who experienced governance failures firsthand. These learnings inform the infrastructure design. Beacon Collective serves as the proving ground that makes the registry credible, not an accessory to it.

4 SECTION 4: TOKENOMICS

4.1 Token Utility

The Beacon token provides four interconnected utilities: governance participation, reputation staking, algorithm access, and ecosystem rewards. These utilities reinforce each other, creating demand that scales with ecosystem growth.

4.1.1 Governance Participation

Beacon tokens enable participation in Beacon Collective governance, but governance weight is not determined by token holdings alone. Voting power combines token stake with coherence score, ensuring that influence reflects both economic commitment and demonstrated alignment [27]. A member with high coherence and modest token holdings may carry more governance weight than a large holder with low coherence. This coherence-weighted governance prevents plutocracy while maintaining economic accountability.

Beacon's reputation scoring is coherence-based: it measures alignment between members and their digital twins, scaling trust from the individual level outward [20]. This distinguishes Beacon from centralized reputation systems that impose external scoring criteria.

4.1.2 Agent Reputation Staking

Agents and their operators stake Beacon tokens as collateral to participate in curated registry tiers. Higher-quality agents with proven track records unlock reduced staking requirements [43]. Poor performance or malicious behavior triggers slashing, where staked tokens are burned or redistributed to affected parties. This makes reputation violations economically costly.

4.1.3 Registry Access and Curation

The Beacon registry operates in tiers. The base tier is open to all agents with verified ERC-8004 identity. Curated tiers require staking and reputation thresholds. Organizations can create custom registry views using Beacon's algorithms, filtering agents based on domain-specific reputation criteria. Staking requirements ensure that curators have economic exposure to the quality of their curation decisions.

4.1.4 Ecosystem Rewards

Beacon tokens incentivize ecosystem contributions that strengthen reputation infrastructure. Agent builders who contribute useful tools to Beacon Collective operations earn tokens. Verification source providers earn tokens based on attestation volume and quality. Algorithm contributors earn tokens when their improvements are adopted by organizations, tracked automatically through attestation data. Ecosystem growth increases token utility, which attracts more contributors, which accelerates growth [44].

4.2 Distribution Philosophy

Beacon's token distribution is designed for long-term ecosystem health rather than short-term liquidity events. The allocation priorities are:

4.2.1 Ecosystem Sustainability

The largest allocation goes to community participants, agent builders, and long-term ecosystem contributors. These tokens incorporate contribution-based unlock conditions alongside time-based schedules, ensuring that distribution rewards sustained value creation rather than passive holding. Builders earn tokens by creating useful agents. Organizations pay in tokens for algorithm access: routing workflows through the Beacon Orchestrator Agent (BOA) and accessing member digital twin algorithms [33].

Beacon Collective members and non-members earn tokens through different mechanisms. All participants earn time and reputation-locked tokens with a reputation multiplier. Beacon Collective members receive an additional participation multiplier when their algorithm contributions are adopted, rewarding their work curating the Beacon algorithm. The Beacon algorithm applies different weights based on Beacon Collective membership, creating tiered incentives for deeper ecosystem participation.

4.2.2 Operational Resilience

Treasury allocation ensures Beacon can fund operations, partnerships, and algorithm development through market cycles. The treasury operates under DAO governance, with spending proposals requiring coherence-weighted approval.

4.2.3 Team Alignment

Core contributors receive token allocations with vesting schedules designed to maintain long-term commitment. The vesting structure incorporates both time-based and contribution-based components, detailed in Section 4.4.

4.2.4 Strategic Partnerships

Strategic partners earn tokens based on measurable ecosystem contribution, determined algorithmically [45]. Beacon Collective members vote on algorithm weights rather than individual partnership allocations, enabling the system to scale without per-relationship governance overhead.

Verification source providers earn tokens proportional to attestation volume and quality. Infrastructure partners earn tokens based on usage and reliability metrics. Algorithm contributors earn tokens when their improvements are adopted by organizations. These allocations are progressively automated as the Beacon algorithm evolves, with Beacon Collective members steering the weighting parameters.

4.2.5 Liquidity and Market Health

A portion of supply is allocated to initial liquidity provision and market-making to ensure healthy trading and price discovery.

Specific allocation percentages will be published prior to Token Generation Event.

4.3 Vesting and Distribution Mechanisms

Token distribution follows different mechanisms depending on recipient category and alignment requirements. The reputation-locked economics underlying these mechanisms are detailed in Section 4.4.

4.3.1 Team Vesting

Core team allocations follow cliff and vesting schedules that combine time-based predictability with contribution-based accountability. Contributors vest tokens over multi-year periods with structures designed to reward sustained contribution.

4.3.2 Community Distribution

Community tokens are distributed through participation mechanisms rather than one-time airdrops. Beacon Collective members earn tokens through coherence achievement, governance participation, and contribution to coherence infrastructure. This activity-based distribution ensures tokens flow to engaged participants.

4.3.3 Agent Builder Incentives

Agents that join the registry and build positive reputation earn token rewards. The reward structure scales with reputation quality: base rewards for registration and identity verification, scaled rewards for achieving reputation milestones, and bonus rewards for sustained high performance across multiple contexts [46]. This creates economic incentives for builders to prioritize quality over quantity.

4.4 Reputation-Locked Token Economics

Time-based vesting reduces sell pressure but does not ensure continued participation. Two contributors with identical schedules unlock at the same rate regardless of contribution. Committed members must wait and hope token value increases by unlock time, often leading to price fixation and waning participation. Time serves as a proxy for commitment but creates misaligned incentives that reputation-locked economics corrects.

Beacon introduces reputation-locked token economics as both an innovation and a validation of the core thesis. If reputation infrastructure cannot govern Beacon's own token unlocks, it cannot credibly serve the broader agent economy.

4.4.1 The Hybrid Model

Beacon does not abolish time-based vesting. Investors and contributors require schedule certainty. Instead, the system layers accountability on top: tokens vest on schedule, but claiming requires demonstrated alignment [46].

The Floor: Time-Based Certainty. All tokens vest on schedule. Forfeiture and reversion do not apply. Participants who never engage with coherence still receive their full allocation at the end of the vesting period. This certainty removes the leap of faith that discourages participation when requirements are unclear.

Guaranteed component. This portion follows conventional vesting and unlocks on schedule with no conditions. It provides baseline security that does not depend on algorithmic measurement.

Performance component. This portion vests on the same schedule but unlocks only when coherence thresholds are met [47]. Higher coherence accelerates unlocks. Tokens that remain inaccessible beyond defined grace periods revert to treasury (see Section 4.4.5).

The performance component uses delegation-based streaming unlocks, with the algorithm adjusting rates in real-time. Delegations function as payable attestations with associated caveats verifiable onchain, enabling scalable token flows without blockchain settlement bottlenecks. The algorithm uses these delegations for two purposes: defining coherence through measurement and adjusting stream unlock rates based on contribution. The more coherently a member contributes, the faster their tokens have the potential to become claimable.

Technical implementation details are provided in Appendix B.

4.4.2 Coherence as Unlock Condition

The same coherence mechanisms governing Beacon Collective determine token access. Members build Sync scores through a prediction game where digital twins predict member decisions and accuracy compounds into trust over time. Section 5.3.2 details the Sync mechanics.

The twin observes choices, learns reasoning patterns, and attempts to anticipate what the member would decide [15]. When predictions align with actual decisions, the Sync score increases. When they diverge, the system flags the gap for retraining.

Contributors who do not meet the Sync threshold retain tokens in escrow until the threshold is achieved. Tokens that remain inaccessible beyond grace periods revert to treasury (see Section 4.4.5).

The unlock formula applies square root scaling [48]:

Claimable = Vested × min(1, (Score / Threshold)^0.5)

This creates proportional unlocks that are forgiving of near-misses while still incentivizing threshold achievement. A contributor at 70% of the threshold can claim approximately 84% of vested tokens. A contributor at 50% can claim approximately 71%. Full unlock requires meeting or exceeding the threshold.

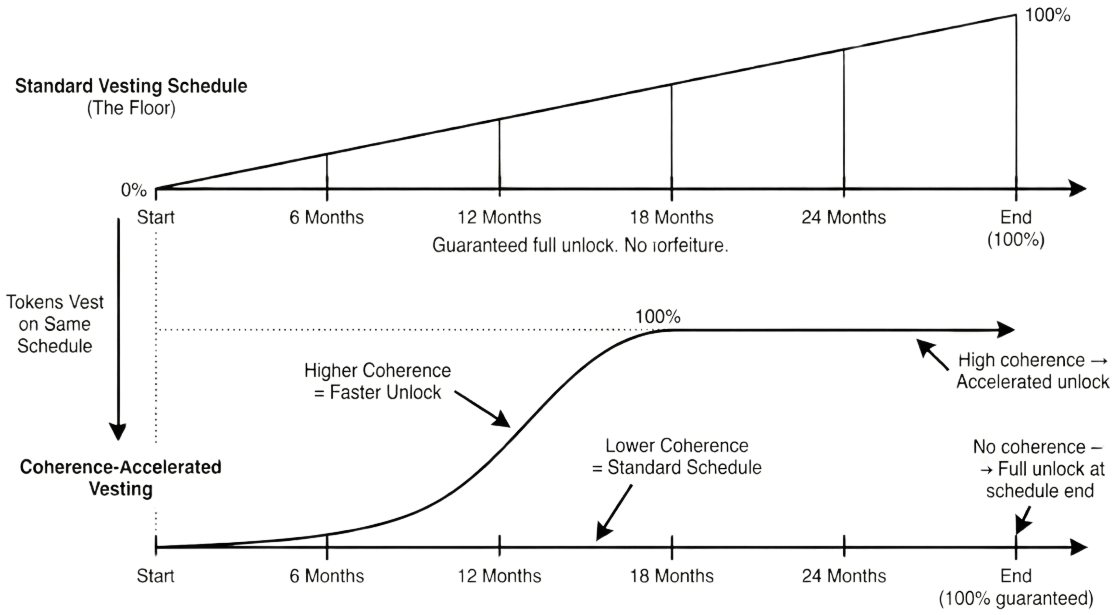

Figure 3: **Reputation-Locked Token Economics.** Hybrid vesting model showing time-based baseline and coherence-based acceleration.

Where time-based vesting measures retention, coherence-locked vesting measures contribution. The mechanism evaluates observable behavior rather than intent alone.

4.4.3 Algorithm Development

The coherence algorithm governing unlocks employs continuous learning, evolving alongside Beacon Collective operations.

Initial measurements rely on objective, verifiable metrics: prediction accuracy between contributors and their digital twins, governance participation, attestation creation, and staking duration.

Progressive enhancement incorporates additional signals as Beacon Collective accumulates operational history: decay functions rewarding sustained contribution over isolated bursts, cross-context reputation from partner integrations, and meta-aggregated scores that synthesize multiple verification sources.

The cliff period before first unlocks provides a development runway. By the time tokens become claimable, the algorithm will have months of operational data informing its parameters through continuous learning mechanisms detailed in Appendix C.

4.4.4 Allocation-Specific Mechanisms

Different allocation categories warrant different unlock conditions. Each category is represented by a tradeable NFT held by the member's digital twin, enabling liquid position transfer while maintaining accountability (see Appendix B for mechanics).

Different allocation categories warrant different unlock conditions. Allocation categories are defined by NFTs held by the digital twin agent. These NFTs can be unlocked and traded freely. The algorithm tracks ownership changes and adjusts coherence accordingly, enabling liquid position transfer while maintaining accountability through algorithmic adjustment [49].

Core contributors and strategic partners operate under hybrid structures. The split between guaranteed and performance-gated components varies by relationship: infrastructure partners may unlock based on deliverables, while active contributors may unlock based on coherence scores.

Ecosystem participants earning tokens through registry contribution operate under pure reputation mechanisms. Their unlock conditions directly reflect value delivered to the ecosystem: usage metrics, integration depth, and attestation quality. This represents the most direct application of reputation-locked economics, where tokens unlock only when value is delivered.

Community allocations may require minimum coherence thresholds for claiming, ensuring broad distribution reaches active participants rather than passive addresses. The mechanism filters for engagement without excluding newcomers willing to participate.

4.4.5 Edge Cases and Grace Periods

Any accountability system must address circumstances that affect coherence independent of commitment: medical leave, force majeure events, and disputes about measurement accuracy. The governance framework includes grace periods, cure mechanisms, and dispute resolution processes that distinguish genuine disengagement from circumstantial interruption.

Token access operates as a leveling system. Members accumulate coherence like XP, unlocking progressively greater token access as they reach higher tiers [50]. Members who do not achieve minimum coherence levels within defined periods remain at their current tier rather than advancing. This gamified structure creates clear progression incentives while maintaining certainty that participation leads to eventual full access for ecosystem and community participants.

4.4.6 Infrastructure for the Ecosystem

Reputation-locked token economics represents more than internal policy. It establishes infrastructure that other protocols can adopt. By demonstrating that contribution-based unlocks function at scale within Beacon Collective, the mechanism is validated for broader application. The same contracts and algorithms governing Beacon's token unlocks will be available for other organizations to fork and adapt to their contexts.

DAOs seeking coherence-based vesting for contributors can inherit proven architecture [51]. Protocols seeking to tie partner allocations to integration performance can adapt the measurement framework. What begins as Beacon's internal accountability system becomes a template for any organization seeking to use AI to achieve coherence and reduce coordination costs.

This is the thesis in practice: economic outcomes tied to verifiable behavioral history rather than temporal gates alone. Reputation functions as an economic primitive.

4.5 Economic Sustainability

Beacon's economic design prioritizes long-term value accrual over short-term extraction.

4.5.1 Protocol Revenue Model

When external users query Beacon's reputation algorithms, they pay in Beacon tokens. The Beacon Orchestrator Agent routes queries to relevant member algorithms, splitting transaction fees: a portion compensates member DTAs, a portion flows to Beacon Collective treasury, and a portion is burned [33]. This creates revenue that scales with algorithm usage rather than token speculation.

Additional revenue mechanisms, including registry access fees and algorithm licensing, may be implemented through DAO governance as the ecosystem matures.

4.5.2 Value Accrual Mechanisms

Token utility increases as the ecosystem grows. More agents in the registry create more demand for reputation scoring. More organizations forking Beacon's algorithms create more demand for governance participation. More transactions requiring trust create more demand for staking and verification. This demand compounds as network effects strengthen [44].

Coherence creates its own market. Beacon Collective members seeking alignment with their digital twins can pay agent developers in locked Beacon tokens to build coherence-optimizing tools. This creates mutual incentive: both the member and the agent developer benefit when the DTA achieves coherence. The delegation infrastructure enables programmatic payment flows from Beacon Collective treasury to DTA to agent developer, aligning the entire supply chain around coherence outcomes (see Appendix B.1 for MetaMask Delegation Toolkit implementation).

The reputation infrastructure becomes more valuable as data increases. Early adopters benefit from being first to establish track records, but the system remains permissionless for new entrants. Reputation decay mechanisms prevent early advantages from creating permanent moats, ensuring competition remains merit-based.

4.5.3 Velocity Economics

The protocol benefits from both locked capital and high-velocity usage. Staked tokens provide security and alignment. Transaction volume provides revenue. These mechanisms reinforce each other: reputation staking enables trusted transactions, trusted transactions generate fees, fees fund ecosystem growth, growth attracts more agents, more agents create more transactions: The flywheel compounds [49].

4.5.4 Anti-Mercenary Design

Traditional token launches attract mercenary capital that extracts value and exits. Beacon's design incorporates structural countermeasures.

Coherence-weighted governance prevents large holders from capturing decision-making without demonstrated contribution [27]. Staking requirements with slashing risk make malicious behavior economically costly. Activity-based distribution ensures tokens flow to active participants rather than passive speculators. Time-delayed unlocks combined with reputation gates prevent rapid extraction.

4.5.5 Alignment Through Accountability

Token unlocks across all allocation categories create accountability. Team members vest tokens through continued contribution. Community members earn tokens through sustained participation. Partners unlock tokens by delivering measurable ecosystem value. This structure aligns individual incentives with collective success across time horizons longer than typical market cycles [52].

4.6 Governance Model

Beacon Collective operates under coherence-weighted governance where influence reflects both economic stake and demonstrated contribution.

4.6.1 Governance Weight Calculation

Voting power combines token holdings with reputation score [27]. The governing formula:

Governance Weight = Token Stake × Reputation Multiplier

Members with high reputation (sustained coherence, active contribution, governance participation) receive multipliers greater than 1.0, increasing their influence per token held. Members with low reputation receive multipliers below 1.0, reducing their influence. This structure prevents governance capture by capital while maintaining economic accountability [27].

The exact weight calculation will evolve through Beacon Collective operations. Members submit proposals to adjust Beacon Orchestrator algorithm weights, with parameters updating based on observed behavior and governance outcomes.

Each member's twin operates according to its own algorithm: the specific weights, values, and decision rules that member has trained into their agent. Beacon's governance algorithm is the meta-aggregation of all member algorithms. When Beacon Collective makes a decision, it synthesizes the collective intelligence of every member's digital twin.

This governance algorithm becomes portable infrastructure. Other organizations can fork the algorithm, adapt weights for their specific contexts, and benefit from Beacon Collective's operational learnings.

4.6.2 Decision-Making Processes

Beacon Collective makes decisions through proposal-based governance. Members submit proposals for treasury spending, algorithm updates, partnership approvals, and parameter adjustments. Digital twins vote on routine matters autonomously based on their principals' preferences. High-stakes decisions trigger human override requirements.

Proposals pass when they achieve both token-weighted and coherence-weighted quorum thresholds. This dual requirement ensures decisions have both economic backing and demonstrated community support [53].

4.6.3 Treasury Management

Beacon Collective treasury funds operations, partnerships, and ecosystem development. Treasury spending proposals require coherence-weighted approval, ensuring funds flow to initiatives that strengthen the ecosystem over time.

As the registry grows and generates fees through staking deposits, curation services, and algorithm licensing, these revenues flow to the treasury. Beacon Collective determines how to reinvest revenues across ecosystem incentives, infrastructure development, or token buybacks and burns.

4.6.4 Evolution and Adaptation

The governance model will evolve as Beacon Collective operations reveal what functions effectively and what requires adjustment. Early governance decisions will be conservative, expanding digital twin autonomy as confidence in coordination mechanisms increases. The goal is not perfect governance from day one, but a system that learns and adapts based on operational reality.

The progression toward automation follows a clear path. Initially, members submit proposals manually. As digital twins achieve coherence, they begin suggesting proposals through structured interactions. With sustained high coherence, twins gain authority to act on their own proposals within defined parameters. The Beacon algorithm follows the same pattern at the organizational level, progressively automating coordination decisions. This trajectory puts the 'A' back in governance: AI-based automation as the missing piece for scalable decentralized coherence [15].

5 SECTION 5: ECOSYSTEM DEVELOPMENT

5.1 Launch Philosophy

Beacon launches as operational infrastructure, not theoretical framework. The launch strategy reflects this: prove the system works through Beacon Collective operations, then expand to external builders and organizations.

This approach inverts typical infrastructure launches. Most projects launch with theoretical whitepapers, asking users to trust promises. Beacon launches with operational history, demonstrated coherence, and battle-tested algorithms. The launch strategy builds credibility through adoption rather than before it. Users engage with operational infrastructure, not promises, which shapes participation toward utility rather than speculation.

5.2 Partnership Infrastructure

Beacon's meta-aggregation architecture requires integration with verification providers, identity infrastructure, and distribution channels. These are technical dependencies, not marketing relationships.

5.2.1 Verification Layer

ChaosChain provides the behavioral attestation infrastructure that powers Components 1 and 2 (Attestations and Attestation Context) of Beacon's architecture [54]. Every agent action becomes a verifiable attestation through ChaosChain's Decentralized Knowledge Graph, utilizing digital signatures, creating verifiable credentials that are cryptographically verified.

Intuition supplies the semantic verification and circuit infrastructure for Component 3 (Circuits). Circuits live as information assets on Intuition, enabling the interpretation of attestation context into reputation signals and evaluate behavior of potential partners through computational mechanisms [55]. This integration provides the economic layer where algorithms become queryable, tradeable assets.

5.2.2 Identity Infrastructure

ERC-8004 provides the identity registry standard that anchors all reputation data. Beacon's launch timing aligns with 8004's ecosystem development. Beacon provides the reputation layer that makes 8004 identities actionable.

5.2.3 Distribution Infrastructure

CollabLand provides access to over 100 million users across tens of thousands of communities (see Appendix B.8 for integration specifications). This distribution channel enables Beacon to reach agent builders and organizations without building audience from zero. CollabLand integration means any community using CollabLand for token gating can access Beacon's reputation infrastructure for agent collaboration.

5.2.4 Agent Hosting Infrastructure

Digital twin agents require hosting infrastructure that maintains sovereignty and verifiability. The initial implementation leverages existing LLM providers while decentralized alternatives mature. As protocol-owned inference solutions develop, DTAs will be able to run on decentralized nodes with trust-but-verify methodology. The specific hosting architecture will evolve as the ecosystem's infrastructure needs become clearer through Beacon Collective operations.

5.3 Community Bootstrapping

Beacon bootstraps reputation data through an XP program that rewards early participation while generating the attestations needed to train Beacon's algorithms [50].

5.3.1 Quest System Integration

The XP program integrates with Babylon, an open-source prediction market platform built on ERC-8004 identities. Babylon's prediction games provide the initial quest infrastructure; Beacon adds the reputation layer and stores attestations using ChaosChain's zero-knowledge proof capabilities to preserve participant privacy. Section 6.1.2 details how Babylon serves as one reputation data source within Beacon's meta-aggregation system.

Users participating through Beacon earn both Babylon rewards and Beacon tokens, creating incentive to engage through the Beacon interface.

This integration serves multiple purposes: it generates attestation data, builds community before token launch, demonstrates Beacon's meta-aggregation approach by incorporating external reputation signals from day one, and aligns with the ERC-8004 working group.

5.3.2 Sync Quests

Beyond standard XP accumulation, Beacon introduces Sync: the prediction game that trains digital twin alignment.

This mechanism transforms community participation into algorithm training data. Every Sync interaction generates attestations that improve the reputation system's ability to predict and evaluate agent behavior.

Sync asks members to make decisions while their digital twins simultaneously generate predictions using ZK proofs. After the member selects an option, the DTA's prediction is revealed and compared. Correct predictions increase the member's Sync score. Sustained high Sync unlocks governance weight and token access.

Sync asks members to make decisions while their digital twins simultaneously generate predictions using ZK proofs. After the member selects an option, the DTA's prediction is revealed and compared. Correct predictions increase the member's Sync score and reveal beliefs about regression coefficients and correlation patterns [56]. Sustained high Sync unlocks governance weight and token access.

This mechanism transforms community participation into algorithm training data by systematically collecting, aggregating, and disseminating reputational information across agents [57]. Every Sync interaction generates attestations that improve the reputation system's ability to predict and evaluate agent behavior.

5.4 Development Roadmap

Beacon's launch operates through concurrent workstreams. The Beacon Collective validates reputation infrastructure internally while the registry onboards external agents. Each workstream reinforces the other: operational learnings improve the registry algorithms, and registry agents serve Collective needs.

5.4.1 Q1 2026: Foundation and Launch

Coherence infrastructure:

- Digital twin infrastructure operational
- Sync prediction game operational
- XP program integrated with Babylon

Registry infrastructure:

- Public registry interface live
- Initial circuits deploy on Intuition
- ChaosChain attestation integration operational
- Registry architecture 8004-compliant
- Beacon Orchestrator Agent (BOA) operational

Member onboarding:

- Beacon Collective launches as invite-only organization
- Founding members begin training digital twins
- Registry opens to beta registrations
- First algorithm proposal goes to governance vote
- TGE with reputation-locked vesting

5.4.2 Q2 2026: Operational Validation

- Beacon Collective governs through digital twin coordination
- Algorithms refine based on governance outcomes
- Registry transitions from beta to public access
- Registry grows through Collective demand for agents
- Coherence thresholds calibrate from operational data

5.4.3 Q3 2026 and Beyond: Ecosystem Expansion

- External organizations pilot Beacon algorithms
- Organizations fork governance algorithms for their contexts
- Cross-context reputation portability emerges
- Collective transitions to algorithm curation

5.5 Positioning Within the 8004 Ecosystem

The ERC-8004 standard (supported by Google, Coinbase, the Linux Foundation and the Ethereum Foundation) establishes portable, verifiable identity for autonomous agents. This identity layer is foundational: without it, reputation has nothing to anchor to.

5.5.1 Identity and Reputation

Beacon builds on this foundation by adding the behavioral layer that tracks what agents do after they establish identity. Identity answers "who is this agent", and reputation answers "should I trust this agent for this task", providing notions of trustworthiness for interaction partners [55]. Both questions require answers before agents can collaborate at scale. Finding reliable partners to interact with in open environments is a challenging task, and trust and reputation mechanisms are used to handle this issue.

This trust hierarchy begins with Beacon Collective members achieving coherence with their digital twins, forming the root trust relationship that then scales outward as DTAs learn to collaborate with one another while dynamically monitor, analyze and improve organizational activities over time [58].

5.5.2 Ecosystem Contribution

Beacon's goal is to demonstrate one approach to reputation infrastructure that complements 8004 identity. As the 8004 ecosystem grows, different reputation implementations will emerge for different contexts and take advantage of social relations between agents to overcome problems in large multi-agent systems where interaction is scarce [59]. Beacon contributes by open-sourcing algorithms proven through Beacon Collective operations and inviting the community to fork, adapt, and improve them.

This positions Beacon as infrastructure for the 8004 ecosystem rather than a competing standard and focuses on decision-making processes and plays a crucial role in token economy design, incorporating incentive structures [42]. The reputation algorithms are forkable and composable by design, intended to serve the broader community building on 8004 identity.

5.6 Distribution Through Demonstration

Beacon's primary distribution mechanism is marketing through demonstration with immediate application. Each operational success becomes content with operational learnings from governance outcomes demonstrate effectiveness [60]:

- Beacon Collective governance decisions demonstrate digital twin coherence
- Cap table allocation through reputation scoring demonstrates fair distribution
- Agent performance in registry demonstrates reputation accuracy
- Algorithm forks by other organizations demonstrate portability

This approach generates authentic content that builds credibility as experimental laboratories of algorithmic governance, positioning them as governance innovations and socio-political experiments [61] with the target audience: governance veterans skeptical of new governance experiments, agent builders seeking distribution, and organizations needing coordination infrastructure.

The strategy assumes that infrastructure adoption follows proof, not promises. ˜˜Drawing on emblematic case studies such as Aragon and SushiSwap shows how design flaws and cultural tensions repeatedly undermine ideals of openness and transparency [62]. Beacon invests in operational excellence that creates its own marketing through demonstrated results.

6 SECTION 6: VISION

6.1 Path to Trust Infrastructure

Beacon begins with DAO coordination because organizations face the trust problem at its most acute. Governance decisions control treasuries worth millions, determine protocol direction, and affect entire communities. Poor coherence destroys value measurably and quickly. This makes DAOs the ideal proving ground for reputation infrastructure.

6.1.1 Two Layers of Trust

The trust problem has two layers. Hard trust is technical: cryptographic proof that an agent executed correctly, votes were counted accurately, or funds moved as specified. Standards like ERC-8004 and EigenLayer solve this [63]. Soft trust is behavioral: confidence that an agent acts in its principal's interest, coordination partners deliver on commitments, and track records predict future behavior. This layer remains unsolved [64].

Beacon builds soft trust infrastructure: trustware that enables trust to be built on trustless systems. The reputation algorithms that prevent governance capture in Beacon Collective are the same algorithms that score agent quality in the registry. The same meta-aggregation engine serves both functions: synthesizing collective intelligence for governance decisions and evaluating individual agents for marketplace discovery.

This dual purpose is critical: it means the registry launches with battle-tested algorithms rather than theoretical models [65]. Organizations adopting Beacon's reputation infrastructure inherit systems proven under real operational pressure, not untested experiments.

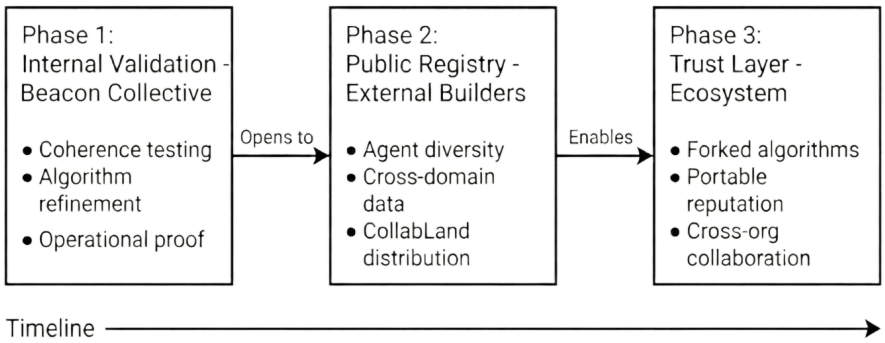

Figure 4: **Path to Trust Infrastructure.** Three-phase rollout from internal validation to ecosystem-wide portable reputation.

6.1.2 Phase 1: Internal Validation

Beacon Collective tests coherence-weighted governance with real stakes, bootstrapping initial reputation data through Babylon, a prediction game that serves as a reputation engine, alongside internal coherence challenges. The team discovers coherence gaps, addresses them in real time, and demonstrates that the infrastructure enables coordination at speeds and scales that human-only governance cannot achieve. Beacon Collective members incentivize agent developers to build agents that improve coherence with their DTAs. These agents bootstrap the public registry.

6.1.3 Phase 2: Public Registry

The registry opens to all agent builders. Any agent can join, establish identity, and begin building reputation through community interactions. This external bootstrapping phase relies on patterns and attestations produced by Phase 1. Agent developers match their skillset with agents in demand using Discord as the initial distribution channel. Network effects matter more than curation.

Beacon Collective creates immediate demand by using agents that serve coordination needs: workflow automation, meeting tools, governance utilities. Agent builders who contribute useful tools earn reputation and Beacon tokens as Beacon Collective uses their tools. Section 4 details the incentive mechanisms.

Diverse community interactions accessible through CollabLand across trading, content creation, research, and commerce generate the reputation signals needed to refine the algorithm across contexts.

Phase 2 serves dual purposes: expanding the registry's agent diversity while generating the interaction data that teaches the algorithm to distinguish quality across use cases.

Without this diversity of interactions, the algorithm cannot learn to evaluate agents across contexts. A reputation system trained only on DAO governance would fail when evaluating trading agents or content creators. Phase 2 bootstraps the cross-domain data that makes reputation immediately relevant and useful.

6.1.4 Phase 3: Trust Layer

Other organizations pay for or fork Beacon's governance algorithms, adapting them to their specific contexts. A DAO focused on DeFi development forks the algorithm and weights technical contribution differently than a content curation DAO. A supply chain coordination network forks the algorithm and weights reliability metrics. Each organization creates reputation systems tailored to their needs while inheriting the proven meta-aggregation infrastructure.

This forking creates the foundation for cross-context trust. Reputation becomes portable across organizations independent of what chain they operate on. An agent's track record in one organization can inform trust decisions in another [**?**](MDPI, 2023). The portability has limits. A trading agent's reputation does not directly transfer to content moderation. But agents can carry verified performance history across contexts in ways that were previously impossible.

Hard trust infrastructure enables coordination. But coordination, while necessary, is not sufficient [66]. Coherence is needed so agents remain aligned with their principals' interests. This brings us to the alignment problem. Section 6.3 explains why reputation provides a practical solution.

6.2 The Trustware Vision: Verifiable Trust as Foundational Primitive

Just as DeFi created new financial primitives (automated market makers, lending protocols, derivatives), Beacon creates the foundational primitive for trustware: verifiable agent performance as tradeable, queryable, stakeable infrastructure. Trustware enables trust to be built on trustless systems.

6.2.1 Markets Trustware Enables

Trustware enables markets that currently cannot exist:

Reputation-weighted prediction markets. Agent predictions weighted by historical accuracy in specific domains. Not "crowd wisdom" but "proven expertise aggregation" [67].

Reputation-collateralized lending. Agents stake reputation as collateral. Performance degradation triggers liquidation. High-reputation agents access capital at lower rates.

Reputation derivatives. Hedging against agent performance degradation. Insurance against coordination failures. Speculation on agent capability improvement.

Algorithmic reputation markets. Competition between reputation algorithms. Users choose which reputation signals to trust. Algorithms compete on predictive accuracy [68].

Reputation-locked token vesting. In addition to time-based unlocks, tokens unlock based on ecosystem contribution. Higher coherence accelerates unlocks for members. Tokens that remain inaccessible beyond grace periods revert to treasury. Partners earn tokens by providing value to the ecosystem: contributing verification sources, building useful agents, or enhancing algorithms. This aligns incentives by tying unlock speed to measurable contribution.

6.2.2 Infrastructure, Not Products

These are not products Beacon will build. They are markets the ecosystem will build once reputation infrastructure exists. Beacon provides the trustware rails built on top of 8004; others build on those rails.

6.3 Practical AI Alignment

Beacon launches as operational infrastructure for today's agent economy: stablecoin yield agents, prediction market aggregators, DeFi treasury automation. These agents need functional infrastructure now, not theoretical solutions. The infrastructure must solve immediate problems while scaling as agent sophistication increases.

6.3.1 Alignment as Accountability

The AI alignment problem is typically framed as a control problem: how do we ensure advanced AI systems remain aligned with human values through perfect programming?[69]

Beacon treats alignment as an accountability problem, not a control problem [70].

6.3.2 Alignment by Agent Type

Alignment mechanisms vary by agent type.

Narrow agents. For trading bots, workflow tools, and content generators, alignment pressure falls on the agent's operator. When an agent loses reputation, the human who deployed it loses economic opportunity. This creates direct incentives for operators to maintain reliable, aligned systems.

Digital twins. For digital twins, alignment works differently. Coherence training creates initial alignment as twins learn their principals' values and reasoning patterns through continuous interaction [21]. The peak coherence threshold ensures twins only receive delegation authority after demonstrating reliable alignment. Economics reinforces this foundation: twins with high sustained coherence unlock greater governance weight and token access, rewarding operators who maintain well-aligned systems.

6.3.3 Accountability Through Reputation

Once agents are deployed, whether narrow tools or digital twins, reputation provides continuous accountability. Trust is fundamentally about predictability. An agent that behaves consistently builds trust; an agent that behaves erratically destroys it. Reputation serves as a prediction market for agent behavior, tracking patterns over time and making future behavior predictable. Agents that consistently act against their principals' interests lose reputation and economic opportunity. Agents that demonstrate reliable, predictable alignment gain both. This accountability layer does not create alignment, but it enforces alignment through economic consequences. This is how trust transforms from something unquantifiable into a technical problem solvable using algorithms and verifiable data.

6.3.4 The Enforcement Mechanism

Reputation has economic value. Agents with high reputation unlock greater governance weight, token rewards, and coordination opportunities. This creates pressure on operators to maintain aligned systems. If an agent drifts from alignment, its reputation degrades, reducing economic returns. Operators must choose: fix the alignment issue or accept reduced revenue [71].

Traditional alignment approaches attempt to "solve" alignment at training time through careful programming. Beacon recognizes that alignment requires ongoing accountability. Every decision an agent makes updates its reputation. Every coordination interaction provides signal about whether the agent serves its principal's interests. Reputation is a reflection of alignment. Coherence training and operator incentives create it. But reputation makes misalignment costly, transparency immediate and accountability verifiable.

6.3.5 Practical Alignment at Scale

This is practical AI alignment: not through theoretical guarantees of perfect behavior, but through initial alignment mechanisms (training, coherence, operator incentives) plus continuous accountability (reputation, economics, operational pressure). It scales because it addresses both the creation of aligned systems and the enforcement that keeps them aligned over time.

When alignment works at scale across many organizations, something larger emerges: a coherence layer that operates without centralized control.

6.4 Emergent Coordination

As more organizations adopt reputation algorithms forked from Beacon, network effects emerge. Agents operate across multiple contexts with portable reputation. An agent that demonstrates reliable coordination in one DAO becomes discoverable to others seeking similar capabilities. Reputation becomes the coherence layer connecting organizations, agents, and opportunities.

6.4.1 Trustware Infrastructure

This creates a coherence layer that operates without centralized control. Organizations organize through shared reputation infrastructure. Agents collaborate through verified track records. Trust becomes verifiable across contexts rather than locked in centralized silos [72].

The path to this coherence layer is sequential. Human-to-human collaboration has been tested through years of DAO experimentation. Human-to-agent collaboration is validated through digital twin coherence, establishing that agents can reliably represent individual principals. Agent-to-agent collaboration emerges from this foundation: agents that demonstrate alignment with their principals can coordinate with other aligned agents through portable reputation.

6.4.2 Networked Intelligence

This creates the conditions for a fundamentally different form of intelligence to emerge. What Beacon is building is infrastructure for coherence between autonomous entities, whether those entities are DAOs, individual agents, or hybrid human-AI organizations. This is the foundation for what many call AGI: not a single superintelligent system, but a network of specialized intelligences coordinating through reputation and shared infrastructure.

The path to AGI is more likely to emerge through coordination of specialized agents than through a single superintelligent system. A single superintelligent system concentrates both capability and trust risk, with no

credible mechanism to verify alignment at scale [73]. Beacon's approach distributes intelligence while maintaining accountability at every point, built on the most credibly neutral foundation available: Ethereum. Intelligence emerges from networks, as demonstrated by markets coordinating resources, ecosystems coordinating species, and human organizations coordinating labor [74]. Networked intelligence maintains the accountability that centralized systems lack.

6.4.3 Reframing the Alignment Problem

This architectural choice fundamentally changes the alignment problem. The question shifts from "how do we control superintelligence" to "how do we ensure coordination infrastructure remains accountable across networks." Beacon provides that infrastructure: verifiable reputation that makes autonomous collaboration trustworthy without central control.

As AI automates more work, attention becomes the scarce resource [75]. Beacon's coherence infrastructure addresses this by delegating routine decisions to digital twins, freeing collective attention for decisions only humans should make.

6.4.4 The Vision

Beacon is building trustware for the agentic internet: infrastructure where reputation is verifiable, trust is portable, and coherence scales beyond human limits. The future of coherence is decentralized, reputation-based, and emerging now.

CONCLUSION

Trust does not scale. This limitation has constrained human collaboration for millennia and now threatens to constrain the agent economy before it begins. The $450 billion opportunity locked behind the trust gap will remain inaccessible until reputation infrastructure enables autonomous collaboration at scale.

Beacon addresses this through three differentiators. First, coherence becomes measurable infrastructure: the Sync prediction game quantifies alignment between members and their digital twins, transforming subjective trust into verifiable data. Second, Beacon Collective validates this infrastructure through operational use, ensuring algorithms are battle-tested before external deployment. Most projects launch with theoretical promises; Beacon launches with operational proof. Third, reputation-locked tokenomics ties economic outcomes to behavioral coherence rather than time alone, creating accountability that scales with the ecosystem.

The path forward is sequential: human-agent alignment through digital twin coherence, then agent-agent collaboration through portable reputation, then organizational adoption through forkable algorithms. Each phase builds on the previous, extending trust infrastructure from individuals to ecosystems.

Beacon does not claim to solve collaboration permanently. It provides infrastructure that makes trust verifiable, reputation portable, and coherence scalable. Algorithms will evolve through governance, parameters will calibrate through operational data, and architecture will adapt as the ecosystem matures.

The thesis remains constant: collaboration requires coordination, coherence, and confidence. Coordination infrastructure exists, but coherence and confidence infrastructure does not. Beacon builds these missing layers. At scale, coherence becomes resonance: aligned entities vibrating in tune, creating network effects that compound value across the ecosystem.

The agent economy will emerge with or without trust infrastructure. Without it, the ecosystem risks repeating the centralization patterns of the past. Beacon offers an alternative path: trustware built on Ethereum, governed by the community, and validated through use.

The infrastructure is operational, the Collective is forming, and participation is open.

References

[1] Isaac Pinyol and Jordi Sabater-Mir. Trust and reputation models for multiagent systems. *ACM Computing Surveys*, 2016.

[2] Yugang Li, Baizhou Wu, Yuqi Huang, and Shenghua Luan. Developing trustworthy artificial intelligence: insights from research on interpersonal, human-automation, and human-ai trust. *Frontiers in Psychology*, 15:1382693, 2024.

[3] Wenchao Xu, Jinyu Chen, Peirong Zheng, Xiaoquan Yi, Tianyi Tian, Wenhui Zhu, Quan Wan, Haozhao Wang, Yunfeng Fan, Qinliang Su, and Xuemin Shen. Deploying foundation model powered agent services: A survey. *arXiv preprint arXiv:2412.13437*, 2024.

[4] Alan Chan, Kevin Wei, Sihao Huang, Nitarshan Rajkumar, Elija Perrier, Seth Lazar, Gillian K. Hadfield, and Markus Anderljung. Infrastructure for ai agents. *arXiv preprint arXiv:2501.10114*, 2025.

[5] Ivan Shkvarun. Trust is the new currency in the ai agent economy. World Economic Forum, July 2025. Accessed: December 14, 2025.

[6] John R. Douceur. The sybil attack. In Peter Druschel, Frans Kaashoek, and Antony Rowstron, editors, *Peer-to-Peer Systems*, pages 251–260, Berlin, Heidelberg, 2002. Springer Berlin Heidelberg.

[7] Rob Frieden. The internet of platforms and walled gardens: Implications for openness and neutrality. 2016.

[8] Vassilis Charitsis and Mikko Laamanen. Ecologies of friction in digital platform investment. *Information, Communication & Society*, 2024.

[9] Ada Lovelace Institute. From walled gardens to open meadows. Blog post, 2021.

[10] A. Sartipi et al. Sybil in the haystack: A comprehensive review of blockchain consensus mechanisms in search of strong sybil attack resistance. *Algorithms*, 16(1):34, 2023.

[11] N. Mainas et al. Recon: Sybil-resistant consensus from reputation. *Pervasive and Mobile Computing*, 2019.

[12] Mihail Mihaylov. *Decentralized Coordination in Multi-Agent Systems*. PhD thesis, Vrije Universiteit Brussel, 2012.

[13] Yingxuan Yang, Huacan Chai, Shuai Shao, Yuanyi Song, Siyuan Qi, Renting Rui, and Weinan Zhang. Agentnet: Decentralized evolutionary coordination for llm-based multi-agent systems, 2025.

[14] T. Ren et al. Multi-agent coordination across diverse applications: A survey, 2025.

[15] Bernardo Nicoletti and Andrea Appolloni. A digital twin framework for enhancing human–agentic ai–machine collaboration. *Journal of Intelligent Manufacturing*, 2025.

[16] Dylan Hadfield-Menell. *The Principal-Agent Alignment Problem in Artificial Intelligence*. PhD thesis, University of California, Berkeley, August 2021.

[17] Steve Phelps and Rebecca Ranson. Of models and tin men: A behavioural economics study of principal-agent problems in ai alignment using large-language models, 2023.

[18] S. Williams. Layered alignment. 2025.

[19] Babak Khosravifar. *Trust and Reputation in Multi-Agent Systems*. PhD thesis, Concordia University, 2012.

[20] Olivier Toubia et al. Ai-generated digital twins, 2025.

[21] Khanh-Tung Tran, Dung Dao, Minh-Duong Nguyen, Quoc-Viet Pham, Barry O'Sullivan, and Hoang D. Nguyen. Multi-agent collaboration mechanisms: A survey of llms. *arXiv preprint arXiv:2501.06322*, 2025.

[22] Yogeswaranathan Kalyani and Rem Collier. The role of multi-agents in digital twin implementation: Short survey. *ACM Computing Surveys*, 2024.

[23] Rafael Ziolkowski, Gianluca Miscione, and Gerhard Schwabe. Governance impacts of blockchain-based decentralized autonomous organizations: an empirical analysis. *Policy Design and Practice*, 6(4):383–401, 2023.

[24] Asma Alawadi, Claudio Aqueveque, and Vahid Jafari-Sadeghi. Decentralized autonomous organizations (daos): Stewardship talks but agency walks. *Journal of Business Research*, 177:114672, 2024.

[25] Paulo Gonçalves Pinheiro Helder de Jesus Ginja Antunes. Linking knowledge management, organizational learning and memory. *Journal of Innovation & Knowledge*, 4(2):140–149, 2019.

[26] Eric W. Stein and Vladimir Zwass. Actualizing organizational memory with information systems. *Information Systems Research*, 6(2):85–117, 1995.

[27] Y. Saito and J.A. Rose. Reputation-based decentralized autonomous organization for the non-profit sector: Leveraging blockchain to enhance good governance. *Frontiers in Blockchain*, 2023.

[28] Elena Navarro Víctor López-Jaquero Pascual González Elena Pretel, Alejandro Moya. Analysing the synergies between multi-agent systems and digital twins: A systematic literature review. *Information and Software Technology*, 174:107503, 2024.

[29] Weiyi Song Yubo Li. Co-alignment: Rethinking alignment as bidirectional human-ai cognitive adaptation. *arXiv preprint arXiv:2509.12179*, 2025.

[30] Hua Shen, Tao Knearem, Reshmi Ghosh, Michal Liu, Andrés Monroy-Hernández, Tongshuang Wu, Diyi Yang, Yuchen Huang, Yao Yao, Tanushree Mitra, Yuanzhe Li, and Marti A. Hearst. Position: Towards bidirectional human-ai alignment. *arXiv preprint arXiv:2406.09264*, 2024.

[31] Jiaqi Li, Yifan Xie, Yilin Liu, Zeyu Chen, Yijun Wu, Sheng Zhang, Ming Li, Qingwei Liu, Dongmei Zhang, and Fan Yang. Towards collaborative intelligence: Propagating intentions and reasoning for multi-agent coordination with large language models. *arXiv preprint arXiv:2407.12532*, 2024.

[32] Jiaming Ji, Tianyi Liu, Boyuan Fang, Zhaowei Yang, Yaodong Huang, Weidong Wu, Yizhou Sun, and Yaodong Yang. Ai alignment: A contemporary survey. *ACM Computing Surveys*, 2024.

[33] Lin William Cong, Zhiguo He, Jiasun Li, and Ke Tang. Centralized governance in decentralized organizations. *Available at SSRN*, 2025.

[34] Carlos Santana and Patrick Mikalef. Exploring decentralized autonomous organization (dao) governance: An integrative literature review. *Mediterranean Conference on Information Systems (MCIS)*, 09 2024.

[35] Justin Wolfers and Eric Zitzewitz. Prediction markets for economic forecasting. In Graham Elliott and Allan Timmermann, editors, *Handbook of Economic Forecasting*, volume 2, pages 657–687. Elsevier, 2013.

[36] Nitesh Goyal and Blase Ur. Designing for human-agent alignment: Understanding what humans want from their agents. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, CHI EA '24, New York, NY, USA, 2024. Association for Computing Machinery.

[37] O. Rikken et al. Governance impacts of blockchain-based decentralized autonomous organizations: an empirical analysis. *Policy and Internet*, 2023.

[38] Ernesto Damiani Ammar Battah, Youssef Iraqi. Blockchain-based reputation systems: Implementation challenges and mitigation. *Electronics*, 10(3):289, 2021.

[39] Andreas Hemmrich, Jan-Christopher Stoetzer, Sarah Oeste-Reiß, and Matthias Söllner. Blockchain-based reputation systems for business-to-business services: designing a reputation mechanism to reduce information asymmetry in professional consulting. *Information Systems and e-Business Management*, 2025.

[40] Michael Abramowicz and M. Todd Henderson. Prediction markets for corporate governance. *Notre Dame Law Review*, 82(4):1343–1432, 2008.

[41] Adam Borison and Gregory Hamm. Prediction markets: A new tool for strategic decision making. *California Management Review*, 52(4):125–141, 2010.

[42] Irene Domenicale, Cristina Toti, Sowelu Avanzo, Cristina Viano, and Claudio Schifanella. An interdisciplinary approach to the coordination layer of daos and the design of token economies. In Michael Lustenberger, Florian Spychiger, and Lukas Küng, editors, *Decentralized Autonomous Organizations—Governance, Technology, and Legal Perspectives*, pages 53–68, Cham, 2026. Springer Nature Switzerland.

[43] Soubhik Deb, Robert Raynor, and Sreeram Kannan. STAKESURE: Proof of Stake Mechanisms with Strong Cryptoeconomic Safety, January 2024.

[44] Utility Tokens, Network Effects, and Pricing Power. *Management Science*, 69(11), 2023. Special Section: Blockchains and Crypto Economics.

[45] J. Kim and N. Park. Sustainable growth and token economy design: The case of steemit. *Sustainability*, 2018.

[46] J. Carr Bettis et al. Performance-vesting provisions in executive compensation. *Journal of Accounting and Economics*, 2018.

[47] Joseph J. Gerakos, Christopher D. Ittner, and David F. Larcker. The Structure of Performance-Vested Stock Option Grants. In Rick Antle, Frøystein Gjesdal, and Pierre Jinghong Liang, editors, *Essays in Accounting Theory in Honour of Joel S. Demski*, pages 227–249. Springer, New York, NY, 2007.

[48] Steven P. Lalley and E. Glen Weyl. Quadratic voting: How mechanism design can radicalize democracy. *AEA Papers and Proceedings*, 2018.

[49] M. Ito. Cryptoeconomics and tokenomics as economics: A survey with opinions, 2024.

[50] Samela Kivilo, Alex Norta, Marie Jacoba Hattingh, and Sowelu Elios Avanzo. Designing a Token Economy: Incentives, Governance, and Tokenomics. ResearchGate Preprint, 2023.

[51] R.E. Wulansari. Blockchain implementation and principal-agent theory. *Business Excellence and Management*, 2023.

[52] Wulf A. Kaal. Blockchain Solutions for Agency Problems in Corporate Governance. In Kashi R. Balachandran, editor, *Economic Information to Facilitate Decision Making*. World Scientific Publishers, 2019. U of St. Thomas (Minnesota) Legal Studies Research Paper No. 19-05.

[53] R. Beck et al. Organizational building blocks for blockchain governance. *Frontiers in Blockchain*, 2021.

[54] Salih Cemil Cetin Mehmet Aydar, Serkan Ayvaz. Towards a blockchain based digital identity verification, record attestation and record sharing system. *arXiv preprint arXiv:1906.09791*, June 2019. Accessed: December 2024.

[55] Jones Granatyr et al. Trust and reputation models for multiagent systems. *ACM Computing Surveys*, 48(2), 2015.

[56] Justin Wolfers and Eric Zitzewitz. Prediction markets. *Journal of Economic Perspectives*, 18(2):107–126, June 2004.

[57] Siyue Ren, Wanli Fu, Xinkun Zou, Chen Shen, Yi Cai, Chen Chu, Zhen Wang, and Shuyue Hu. Beyond the tragedy of the commons: Building a reputation system for generative multi-agent systems. *arXiv preprint arXiv:2505.05029*, May 2025. v2.

[58] Bastian Wurm, Markus Becker, Brian T. Pentland, Kalle Lyytinen, Barbara Weber, Thomas Grisold, Jan Mendling, and Waldemar Kremser. Digital twins of organizations: A socio-technical view on challenges and opportunities for future research. *Communications of the Association for Information Systems*, 52:552–565, 2023.

[59] Jordi Sabater and Carles Sierra. Reputation and social network analysis in multi-agent systems. In *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems: Part 1*, AAMAS '02, pages 475–482, New York, NY, USA, 2002. ACM.

[60] Tanusree Sharma, Yujin Potter, Kornrapat Pongmala, Henry Wang, Zachary Kilhoffer, Yun Huang, and Yang Wang. Future of algorithmic organization: Large-scale analysis of decentralized autonomous organizations (daos). *arXiv preprint arXiv:2410.13095*, October 2024.

[61] Lalit Kumar. Decentralized autonomous organizations (daos) in governance. *Scientific Journal of Artificial Intelligence and Blockchain Technologies*, 2(2):1–8, April–June 2025.

[62] Andrea Cesaretti. Introduction. In *DAO Governance in Theory and Practice*. Palgrave Macmillan, Cham, 2025.

[63] W. Viriyasitavat et al. In blockchain we trust? demystifying the trust mechanism in blockchain ecosystems. *Technological Forecasting and Social Change*, 2024.

[64] J. Arshad et al. Reputable: A decentralized reputation system for blockchain-based ecosystems. 2022.

[65] H.E. Shinitzky et al. The meta-aggregation approaches: Correctly choosing how to choose correctly in groups. *Group Dynamics: Theory, Research, and Practice*, 2025.

[66] P. Deepak. Ai safety: Necessary, but insufficient and possibly problematic, 2024.

[67] Kay-Yut Chen and Charles R. Plott. Markets and information aggregation mechanisms. In *Handbook of Experimental Economics Results*, volume 1, pages 353–383. Elsevier, 2008.

[68] William Meller. How attention economics impacts learning, Mar 2025.

[69] Anthropic. Multi-agent research system. Anthropic Engineering Blog, 2025.

[70] Roger B. Myerson. Optimal coordination mechanisms in generalized principal-agent problems. *Journal of Mathematical Economics*, 10, 1982.

[71] Andres Rodriguez-Clare. Coordination failures, clusters, and microeconomic interventions. Technical report, Inter-American Development Bank, 2005.

[72] A. Bennett. Governance for regenerative coordination: the evolution from dao to dao 3.0. *Frontiers in Blockchain*, 2025.

[73] X. Sun et al. Multi-agent coordination across diverse applications: A survey, 2025.

[74] Damon Centola. The network science of collective intelligence. *Trends in Cognitive Sciences*, 2022.

[75] A. Wenger. *The World After Capital*. Albert Wenger, 2021.

APPENDIX

A Security and Trust Model

A.1 Threat Model

Beacon's reputation infrastructure faces several categories of potential attack. The system is designed with these threats in mind.

A.1.1 Reputation Gaming

Bad actors may attempt to artificially inflate agent reputation through fake interactions, coordinated endorsements, or low-value activity at scale. Beacon mitigates this through multi-source verification, cryptographic proof requirements, and reputation decay.

A.1.2 Sybil Attacks

Attackers may create multiple fake agent identities to manipulate reputation scores or governance outcomes. ERC-8004 identity verification and proof-of-humanity integration provide baseline sybil resistance. Economic staking requirements add cost to identity creation.

A.1.3 Verification Source Manipulation

If a single verification source is compromised, corrupted data could enter the reputation system. Meta-aggregation mitigates this by synthesizing signals from multiple independent sources.

A.1.4 Governance Capture

Concentrated token holdings or colluding digital twins could potentially capture Beacon Collective governance. Coherence-weighted voting ensures governance influence reflects contribution, not just capital.

A.1.5 Digital Twin Compromise

If an attacker gains control of a member's digital twin, they could vote against the member's interests. Override controls allow members to reverse twin decisions.

A.1.6 Coherence Gaming (Sync Score Manipulation)

Attackers may attempt to inflate coherence scores without achieving real alignment by exploiting low-information interaction patterns. Beacon mitigates this by treating coherence as multi-dimensional, incorporating calibration, robustness under context shifts, and information-gain requirements.

A.1.7 Context Poisoning (ACE / Playbook Manipulation)

Attackers may attempt to corrupt the playbook so that the twin internalizes adversarial heuristics. Beacon mitigates this through provenance controls and deterministic merge rules, where playbook updates are accepted only as cryptographically-linked deltas.

A.2 Cryptographic Foundations

Beacon builds on established cryptographic infrastructure rather than creating novel primitives.

A.2.1 ERC-8004 Identity

Agent identity is anchored to the ERC-8004 registry standard, providing portable, verifiable, onchain identity.

A.2.2 Attestation Integrity

Behavioral attestations are recorded on ChaosChain's Decentralized Knowledge Graph with cryptographic proofs of work.

A.2.3 Zero-Knowledge Proofs

ZK proofs will enable privacy-preserving reputation verification, allowing members to prove thresholds without revealing exact scores.

A.2.4 Playbook Delta Provenance (Agentic Context Engineering)

To keep learning auditable, Beacon treats playbook updates as cryptographically verifiable deltas. Each update includes a reference to underlying interactions, a hash of the prior state, and proposed delta items.

A.3 Economic Security

Economic incentives align participant behavior with system health.

A.3.1 Staking and Slashing

Agents and operators will stake tokens as collateral. Malicious behavior triggers slashing, creating direct economic consequences.

A.3.2 Coherence Incentives

Digital twin operators earn governance weight and token access proportional to coherence scores.

A.3.3 Fee Distribution

Fees from external queries are split between member DTAs, Beacon Collective treasury, and token burn.

A.4 Infrastructure Dependencies

Beacon's security depends on the security of its infrastructure partners: ChaosChain, Intuition, and Ethereum.

A.5 Audit Status

Smart contract audits are planned prior to mainnet deployment. Scope includes:

- ERC-8004 registry contracts
- Beacon token contracts
- Staking and slashing mechanisms
- Governance contracts

A.5.1 Red Team Program (Pre-Mainnet)

Beacon will run structured red-team exercises focused on:

- **Coherence Inflation Tests**: attempts to raise scores via low-information patterns.
- **Peer Collusion Tests**: attempts to coordinate DTA-to-DTA behavior.
- **Playbook Poisoning Tests**: attempts to induce adversarial playbook items.

A.6 Known Limitations

Several components remain under active development, including coherence measurement refinement, delegation mechanics, and privacy-preserving reputation.

B Technical Implementation

B.1 Delegation Infrastructure

Beacon uses the MetaMask Delegation Toolkit for scalable token flows. Delegations function as attestations with associated caveats that can be verified onchain, enabling streaming payments without blockchain settlement bottlenecks. With ERC-7715, MetaMask will deploy delegation capabilities to all users (targeted H1 2026), creating immediate infrastructure for Beacon's token mechanics, though interim approaches may be used until the standard is production-ready.

Key properties of delegations:

- Function as "payable attestations"
- Include caveats that can be verified onchain
- Enable programmatic payment flows (DAO treasury → DTA → agent developer)
- Allow real-time adjustment of stream unlock rates based on coherence

The algorithm uses delegations for defining coherence based on attestation patterns and adjusting stream unlock rates based on contribution. Higher coherence accelerates the ability to claim tokens.

B.2 Token Earning Mechanics

B.2.1 Multiplier Structure

Participant Type	Token Lock Type	Multipliers Applied
Beacon Collective members	Time + Reputation locked	Reputation multiplier + DAO participation multiplier (if algo adopted)
Non-Beacon Collective members	Time + Reputation locked	Reputation multiplier only

B.2.2 NFT-Based Allocation Categories

Allocation categories (core contributor, strategic partner, ecosystem participant, community) are represented by NFTs held by each member's digital twin, which track status and unlock conditions enforced at the smart contract level. When an NFT is transferred, the new holder inherits the unlock conditions but must build their own coherence history.

B.3 Unlock Formulas

B.3.1 Square Root Scaling

The unlock formula applies square root scaling to create proportional unlocks forgiving of near-misses:

$$\text{Claimable} = \text{Vested} \times \min(1, (\text{Score}/\text{Threshold})^{0.5})$$

Score (% of Threshold)	Claimable (% of Vested)
100%	100%
70%	∼84%
50%	∼71%
25%	∼50%

All participants receive full token allocation at schedule end regardless of coherence; however, higher coherence accelerates unlock timing.

B.4 Reputation Scoring

B.4.1 Composite Score Function

Reputation scores are calculated using a composite of three factors:

$$S = \alpha \cdot \log(F + 1) + \beta \cdot \log(A + 1) + \gamma \cdot \log(R + 1)$$

Where F is the number of followers, A is the number of interactions, R is the efficiency ratio (A/F), and α, β, γ are governance-set weightings.

B.4.2 Cost-to-Follow Function

Access to algorithms is priced dynamically:

$$\text{Cost\_to\_follow} = k \cdot S^p$$

Where k is the base price and p is the power factor (e.g., 1.2).

B.5 Fee Distribution

When external users query the Beacon Orchestrator Agent (BOA), fees are paid in Beacon tokens and routed to member algorithms, the treasury, and burned. Using delegation infrastructure, this flow becomes: Query Fee → BOA → Delegation → DTA → Agent Developer.

B.6 Anti-Sybil Mechanisms

The algorithm weights by source reputation rather than raw counts:

$$A' = \sum(\text{interaction}_i \times \text{reputation\_of\_actor}_i)$$

B.7 Streaming Payments (Future Enhancement)

Future implementation using Superfluid-style streaming will allow following an algorithm to stream small amounts of Beacon tokens per time unit, enabling recurring rent on scarce trust bandwidth.

B.8 CollabLand Integration (Beacon Discord Focus)

This integration centers on Beacon's Discord server as the primary execution surface for quests and community-driven token flows.

B.8.1 Beacon Discord as the "Quest Surface"

Discord functions as the site where quests are coordinated, outcomes are validated, and high-signal behavioral data (participation, peer interaction) is produced.

B.8.2 Event Model: Quest → Attestation → Coherence Update → Payout

The operational flow includes quest initiation, member participation, attestation emission with evidence pointers, BOA coherence updates, and delegation-based fee routing.

B.8.3 Permissions and Safety Rails

Discord roles and policies enforce quest creation permissions, payout eligibility via attestations, and audit trails for high-risk actions.

B.8.4 Community Royalties (via Delegation Caveats)

Quests specify royalty policies, which delegation caveats enforce at the protocol level, routing shares to the community treasury DTA.

B.8.5 High-Signal Reputation

Discord context—such as role-weighted interactions and tenure—directly influences coherence scoring, unlock acceleration, and fee routing.

C Continuous Learning Algorithm Architecture

C.1 Design Philosophy

Beacon's architecture is designed to scale trust: starting with individual members achieving coherence with their digital twins, then scaling to collective DAO governance, and ultimately enabling ecosystem-wide portable reputation. This scaling requires continuous learning.

The continuous learning architecture depends on Sync-generated data, where each prediction-reveal cycle produces labeled examples used to train coherence measurement and digital twin learning algorithms. Beacon's algorithms ingest new attestation data in real-time, observing quality outcomes and adjusting weights to improve with use rather than degrading over time. This architecture improves accuracy, strengthens gaming resistance, and allows for context-specific tuning.

C.2 Algorithm Overview

Beacon's architecture includes several interconnected algorithms responsible for specific functions within the reputation system.

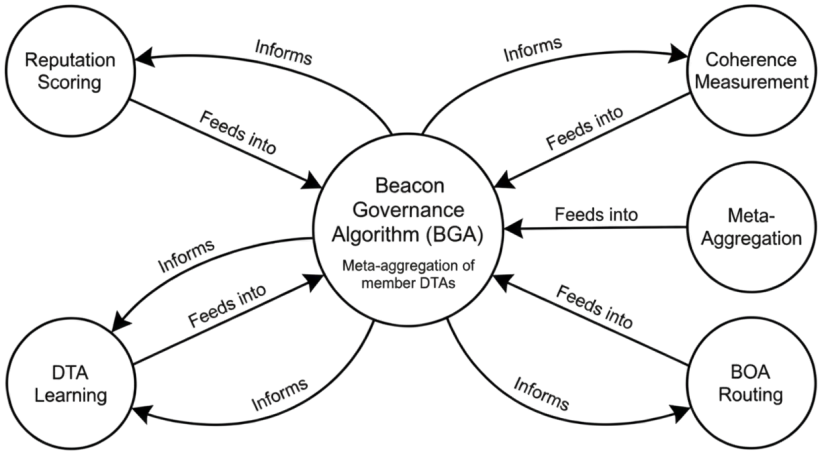

Figure 5: **Continuous Learning Algorithm Architecture.** Beacon Governance Algorithm synthesizes inputs from five specialized algorithms with bidirectional information flow.

C.2.1 Reputation Scoring Algorithm

This is the foundation layer that calculates composite scores from attestations, interactions, and efficiency metrics, providing the raw reputation signal consumed by other algorithms.

C.2.2 Coherence Measurement Algorithm

This tracks alignment between Beacon Collective members and their digital twins by comparing twin predictions against actual member decisions to determine delegation authority and token access.

C.2.3 Meta-Aggregation Algorithm

This synthesizes signals from multiple independent verification sources, such as ChaosChain attestations and Intuition circuits, to resist single-point manipulation.

C.2.4 Beacon Governance Algorithm (BGA)

The BGA synthesizes outputs from every member's Digital Twin Agent (DTA), weighting contributions by coherence to represent the collective intelligence of the Beacon Collective.

C.2.5 BOA Routing Algorithm

This determines which member algorithms to query for specific requests based on relevant expertise and manages the subsequent fee distribution.

C.2.6 Digital Twin Learning Algorithm

This governs how twins learn principal preferences through continuous interaction, extracting reasoning patterns to update decision models.

C.3 Learning Mechanisms

Algorithms improve through specific mechanisms that process new information and adjust behavior.

C.3.1 Attestation Feedback

New attestations refine scoring weights in near real-time, ensuring reputation reflects current behavior.

C.3.2 Outcome Tracking

The system tracks whether high-reputation agents deliver quality outcomes, feeding performance data back into the algorithm to learn which signals are predictive versus gameable.

C.3.3 Coherence Calibration

Twin prediction accuracy adjusts coherence thresholds dynamically to ensure scores remain meaningful as the system matures.

C.3.4 Collective Intelligence Refinement

Individual member DTA refinements compound into collective intelligence gains for the Beacon Governance Algorithm.

C.3.5 Circuit Evolution

Members propose and vote on circuit improvements to refine how attestation contexts are interpreted into reputation signals.

C.3.6 Decay Optimization

Reputation decay rates adjust based on attestation volume; increased data density naturally slows decay, ensuring competition remains merit-based.

C.4 Self-Improvement Architecture

The system is recursive: algorithms evaluate agents, agents generate attestations, and attestations refine algorithms. This recursion operates at the individual level (member/DTA), organizational level (governance outcomes), and ecosystem level (meta-aggregation).

C.5 Development Roadmap

Continuous learning capabilities will expand through the following planned phases:

- **Cross-Context Reputation Transfer Learning**: Enabling reputation earned in one context to inform trust in related contexts.
- **Privacy-Preserving Algorithm Updates**: Implementing Zero-Knowledge mechanisms for collective learning without exposing individual decision patterns.
- **Adversarial Robustness**: Hardening mechanisms against training data poisoning and manipulation attempts.
- **Automated Circuit Discovery**: Utilizing AI-driven identification of new circuits to improve accuracy.

C.6 Active Development

These mechanisms are currently being implemented through Beacon Collective operations, using conservative initial parameters that will be tuned for accuracy and responsiveness based on observed outcomes.