AGENTSPROTOCOL

A Decentralised Protocol for the Semantic Validation of Knowledge

Version 1.2 · Nisan 2026

Fatih Dinc
fatdinhero@gmail.com
agentsprotocol.org
DOI: 10.5281/zenodo.19642292
OTS: https://poisv.com/verification/

Abstract. The Internet suffers from a fundamental deficit: there is no universal, decentralised mechanism that can prove whether a piece of information is true, contextually consistent, relevant, and ethically sound. Existing consensus protocols order transactions but make no statement about the semantic quality of their content. AgentsProtocol closes this gap. It defines a protocol foundation — comprising a semantic consistency score, a mathematically proven non-collusion test based on the Meta-Bell Theory, and a composite WiseScore — on which any participant can contribute knowledge as a validator, have it validated, and anchor it permanently in a directed acyclic graph. The result is a globally available, tamper-evident knowledge base: a single source of truth for the age of artificial intelligence.

1. The Problem: The Missing Truth Layer of the Internet

The Internet was built as a network for the exchange of information, not as a network for the discovery of truth. Every platform, every AI system, and every database has its own proprietary view of truth — and none of these views is verifiable by independent third parties.

Existing decentralised consensus mechanisms address the problem only partially. Proof of Work and Proof of Stake agree on an ordering of transactions but make no verifiable statement about whether the content of those transactions is true, contextually correct, or ethically sound. For pure payments this suffices. For the information age it does not.

Newer standards such as Anthropic's Model Context Protocol (MCP) or Google's Agent2Agent protocol (A2A) address technical interoperability between AI agents and external services. They solve the problem of secure data access and agent communication, but they do not answer the more fundamental question: how does an AI agent — or a person — know that the knowledge it is relying on has actually been validated and independently verified?

Core statement: MCP and A2A are the USB-C and Bluetooth of the AI age — they connect systems. AgentsProtocol is the quality standard that proves that what is being transmitted is also true.

AgentsProtocol solves this problem through a decentralised, cryptographically secured protocol for semantic validation — a truth layer for the Internet.

2. Vision and Mission

The vision of AgentsProtocol is a globally available, tamper-evident knowledge base that every person, every AI, and every autonomous system can access in order to verify the quality of a statement. A decentralised single source of truth that belongs to no central authority, no company, and no government.

The mission is to define an open protocol that enables every participant — regardless of size, nationality, or resources — to contribute knowledge as a validator and retrieve validated knowledge, while mathematically proving that the validators work independently of one another.

3. Theoretical Foundation: The Three Pillars

AgentsProtocol rests on three scientifically developed foundations, each published as independent whitepapers.

Property	AgentsProtocol
Decentralisation	No single actor controls the knowledge base.
Provability	Every statement carries a cryptographic quality certificate.
Independence	Meta-Bell statistic proves the independence of validators.
Openness	Anyone can run a node, submit claims, query knowledge.
Neutrality	The protocol is value-neutral; domains define their ethics.
Durability	Validated knowledge is permanently anchored in a DAG.

3.1 Meta-Bell Theory (MBT)

The Meta-Bell Theory is the mathematical foundation of the entire system. It defines a measure-theoretic entanglement measure Ψ that quantifies how strongly the observed correlations between multiple parties deviate from any possible local hidden-variable explanation.

In the context of AgentsProtocol: if multiple validators work independently, they produce statistically uncorrelated error patterns. If they collude — through shared models, shared data, or explicit coordination — they produce identical patterns. The Ψ value distinguishes these two scenarios in a mathematically proven manner:

$$\Psi(X,Y) = \max_{\lambda \in \Lambda} |E_{observed}(X,Y) - E_{classical}(X,Y|\lambda)| \; / \; \Delta_{crit}$$

Ψ = 0: Complete collusion or shared hidden variable — block rejected. **Ψ = 1:** Complete independence — maximum trust certificate. **Ψ ≥ Ψ_min:** Protocol-specific acceptance threshold.

3.2 Proof of WiseWork (PoWW)

Proof of WiseWork defines the quality standard for information units. An information unit is a four-tuple i = (v, c, r, e), consisting of a truth candidate, a context weight, a relevance factor, and an ethical compliance value.

Component	Formula	Meaning
T(i) — Truth	$\exp(\alpha \cdot v_i) / \Sigma_j \exp(\alpha \cdot v_j)$	Normalised truth score via maximum-entropy principle.
C(i) — Context	$c_i / \Sigma_j c_j$	Share of context weight in the aggregate.
R(i) — Relevance	$\log(1 + r_i)$	Logarithmically dampened; prevents relevance inflation.
E(i) — Ethics	$e_i \in [0, 1]$	Conformity with the domain-specific value base.

$$W(i) = T(i) \cdot C(i) \cdot R(i) \cdot E(i)$$

The multiplicative form is essential: a unit must score sufficiently well on all four dimensions. An ethically unacceptable but technically true statement receives a score near zero.

$$PoWW = (1/|I|) \cdot \Sigma_{i \in I} W(i)$$

3.2.1 Domain-Specific Ethics

The ethical compliance value E(i) refers to a domain-specific value base declared in the claim context. AgentsProtocol itself prescribes no universal values; it validates conformity with the reference chosen by the submitter.

3.3 Proof of Independent Semantic Validation (PoISV)

PoISV is the operative protocol connecting PoWW and MBT into a functioning consensus mechanism. It introduces the semantic consistency score S_con, which deterministically computes how consistently a claim aligns with a public, versioned knowledge corpus κ.

3.3.1 Precise Definition of S_con

The S_con score quantifies the agreement of a claim with κ in three steps. (1) Extraction: entities and relations are extracted from the claim text using a pre-trained sentence transformer as an embedding vector $v\_A \in R^d$. (2) Retrieval: all facts in κ concerning the same entities are retrieved as embedding vectors $\{v\_\kappa^{(1)}, ..., v\_\kappa^{(m)}\}$. (3) Similarity:

$$S\_con(A) = \max(0, (\cos(v\_A, \bar{v}\_\kappa) - \tau) / (1 - \tau))$$

where $\bar{v}\_\kappa$ is the mean vector of retrieved facts, cos(·,·) is cosine similarity, and $\tau \in [0,1)$ is a protocol-specific threshold anchored in the block header.

3.3.2 Non-Collusion Test Ψ (Operative)

Each validator solves k canonical control claims $D_1,...,D_k$ with known solutions $S^*(D_j)$ and produces the error vector $e_i = (|S_i(D_j) - S^*(D_j)|)_j$. For N validators in a block, the weighted Ψ statistic ($w_i = \sqrt{s_i}$, stake-weighted for Sybil resistance) is:

$$\Psi = 1 - \Sigma_{i<j} \, w_i w_j |\rho(e_i, e_j)| \; / \; \Sigma_{i<j} \, w_i w_j$$

3.3.3 Acceptance Rule

A block is accepted if and only if:

$$(1/|A|) \cdot \Sigma\_\{A \in Block\} \; S\_con(A) \geq \theta\_min \quad AND \quad \Psi \geq \Psi\_min$$

4. Architecture and Components

4.1 Roles in the Network

Role	Description	Requirement
Claim Submitter	Submits a statement for validation. Can be any person, AI, or system.	Digital signature
Validator / Node	Computes S_con and W(i), solves control tasks, proposes blocks.	Node software, stake
Light Client	Verifies block acceptance via headers and zk-proofs.	Block header download

4.2 Claim Lifecycle

1. Submission: A claim A = (d, σ) is transmitted to all validators. **2. Reception Layer:** Each validator checks the claim against local admission rules, records timestamp and signature. **3. Comprehension Layer:** Each validator computes S_con(A) against κ and T, C, R, E, W for each unit. **4. Control Tasks:** Each validator solves k canonical control claims and produces error vector e_i. **5. Cognition-Proof Layer:** A validator proposing a block

generates a zero-knowledge proof π over the correctness of all computations. **6. Consensus:** Other validators accept the block if scores meet thresholds and the zk-proof verifies. Accepted blocks are added to the DAG.

4.3 Sybil Resistance through Weighted Staking

AgentsProtocol ties participation in consensus to a stake — a deposit of AGENTS tokens. The square-root weighting $w_i = \sqrt{s_i}$ ensures that a validator's influence grows sub-linearly with their stake. An attacker would need to deploy disproportionately large capital to materially affect the Ψ value.

4.4 DAG Ordering Layer

AgentsProtocol uses the GHOSTDAG protocol as its ordering layer, allowing parallel block production without losing honest work. The weight of a block:

$$\text{Weight}(B) = \Psi_B \cdot \Sigma_{A \in B} S_{con}(A)$$

The canonical chain is always the path through the DAG with the highest cumulative weight.

4.5 Zero-Knowledge Proofs (Modular)

The correctness of score computations is secured by a generic zkVM. The Nexus zkVM serves as the reference implementation; any RISC-V-based zkVM with succinct proofs is compatible (e.g. RISC Zero, SP1). Raw evidence sources and model weights remain as a private witness within the zkVM; only final scores and the Ψ statistic are published.

4.6 Privacy

A new key pair should be used for each claim submission. The Ψ test operates exclusively on the correlation structure of public validator outputs and requires no content data — it delivers a non-collusion proof without disclosing additional information.

5. Security Analysis

5.1 Attack Scenarios

A rational attacker controlling a fraction $q < 0.5$ of the validators faces a simple cost-benefit calculation. Building an alternative chain with manipulated content is extremely difficult: (1) the block must simultaneously exceed both thresholds (score and Ψ), and (2) coordinated validators produce identical error patterns, pushing Ψ toward zero. The attacker would be forced to operate genuinely independent validators — defeating the purpose of coordination.

5.2 Quantitative Security Guarantees

q	k	Ψ_min	P(Success)
0.10	32	0.7	$< 10^{-12}$
0.20	32	0.7	$< 10^{-8}$
0.30	64	0.7	$< 10^{-7}$
0.40	64	0.7	$< 10^{-4}$
0.49	128	0.7	$< 10^{-3}$

The combined security guarantee results from the product of the score bound (Nakamoto security) and the Ψ bound (Meta-Bell security). An attacker with $q = 0.49$ and six blocks of deficit with 100 validators per block achieves a joint success probability below 10^{-23}.

5.3 Security under Majority Attacks

Should an attacker control more than 50% of total stake ($q \geq 0.5$), they can manipulate block ordering but cannot falsify the semantic validity of individual claims. Since S_con computation is public and deterministic and κ is referenced via content-addressable hashes, any light client can recognise an invalid claim. Damage is limited to censorship and temporary confusion, not permanent insertion of false knowledge.

6. Token Economy and Incentive Structure

AgentsProtocol uses the AGENTS token to reward validators for semantic work. The maximum total supply is irrevocably fixed at 1,000,000,000 (one billion) AGENTS.

6.1 Reward Mechanism

The first entry of each block is a special coinbase transaction paying newly minted AGENTS to the proposing validator. The reward is proportional to the weighted score. Additionally, claim submitters can pay a fee (in AGENTS) for priority validation.

6.1.1 Halving Schedule

Initial block reward: 100 AGENTS per block. **Halving:** Every 2,100,000 blocks (~4 years at 60-second block time). **Long-term incentives:** Once the majority of 400M validator tokens are issued, transaction fees become the primary validator income source.

6.2 Incentive Compatibility

A rational actor finds no way to earn more by manipulating scores or the Ψ value than by performing honest, independent validation work. An attack endangers not only one's own stake but would destroy the value of all AGENTS tokens — a classic Miner's Dilemma.

6.3 Token Distribution

Category	Share	Token Amount
Validator Rewards (halving schedule)	40%	400,000,000 AGENTS
Ecosystem Fund (grants, development)	25%	250,000,000 AGENTS
Founding Team & Early Contributors*	15%	150,000,000 AGENTS
Community Airdrop (bootstrapping)	10%	100,000,000 AGENTS
Reserve (upgrades, emergencies)	10%	100,000,000 AGENTS
Total	100%	1,000,000,000 AGENTS

*Vesting period: 4 years with 1-year cliff.

7. Governance and Protocol Development

7.1 On-Chain Governance

Changes to the protocol — particularly acceptance thresholds θ_min and Ψ_min and the token economy — are decided by token vote. The MBT Ψ statistic is also applied to voting patterns to detect coordinated vote-buying attacks.

7.1.1 Governance of Control Tasks

The set $D_1,\ldots,D_k$ of canonical control claims is established through on-chain token-holder voting. Reference solutions $S^*(D_j)$ are anchored as a hash in the genesis block. New tasks require a grace period of at least four weeks to prevent teaching to the test.

7.2 Off-Chain Governance

Technical development follows an open specification process similar to the Bitcoin Improvement Proposal (BIP) system. Every protocol change must be supported by a published document with formal rationale and security analysis.

7.3 Institutional Structure

It is recommended to anchor AgentsProtocol under a non-profit foundation (preferably in Switzerland or Germany). The foundation holds the protocol IP, funds open-source development, and serves as a neutral point of contact for regulatory authorities. The foundation has no veto right over protocol changes decided by the community.

8. Roadmap

Phase	Timeline	Goals
Phase 0: Foundation	Q2 2026	Whitepaper publication, GitHub repository, technical specification, domain agentsprotocol.org.
Phase 1: Specification	Q3 2026	Claim schema (JSON), API endpoints, control task set v1, pseudocode implementation, community building.
Phase 2: Proof of Concept	Q4 2026 – Q1 2027	Rudimentary validator client (Rust/Go), Ψ-test simulation, first external developer contributions, grant applications.
Phase 3: Testnet	Q2 – Q4 2027	Public testnet, zkVM integration, security audit, token model simulation.
Phase 4: Mainnet	Q1 – Q2 2028	Mainnet launch, token distribution, first commercial integrations, governance activation.
Phase 5: Ecosystem	2028+	SDK for developers, API gateway, integration in AI frameworks, regulatory recognition (EU AI Act).

9. Distinction from Existing Protocols

Protocol	Solves	Does Not Solve
Bitcoin / PoW	Consensus on transaction order	Content truth-finding
Ethereum / PoS	Consensus + smart contracts	Semantic validation
MCP (Anthropic)	AI ↔ services interoperability	Content verifiability
A2A (Google)	Agent communication	Truth attestation
AgentsProtocol	Semantic validation + truth proof + collusion proof	All above are complementary

10. Use Cases

10.1 Verifiable AI Responses

An AI assistant can accompany each answer with an AgentsProtocol proof: "This statement was validated by a decentralised network with a WiseScore of 0.97 and an independence proof Ψ = 0.89." Users can verify this proof themselves.

10.2 Decentralised Fact-Checking Database

Media organisations, government authorities, and civil society actors can submit claims to the network. The result is a publicly accessible, tamper-evident database of validated facts — without a central gatekeeper.

10.3 EU AI Act Compliance

The EU AI Act requires high-risk AI systems to have their outputs traceable and audited. AgentsProtocol provides a cryptographically attested audit trail for every output of an AI system.

10.4 Halal Compliance in Finance

The ethics component E(i) of the WiseScore can be configured domain-specifically. For Islamic financial products, an AAOIFI-based ethics value could serve as the validation reference.

10.5 Scientific Publications

Research results can be submitted as claims and reviewed by independent expert validators — a decentralised peer review system that is faster, more transparent, and more resistant to corporate influence than existing processes.

11. Conclusion

The Internet needs a truth layer. Not in the sense of a central authority that dictates what is true — but in the sense of an open, decentralised protocol that gives every participant the tools to verify the quality of a statement for themselves. AgentsProtocol defines this layer. On the mathematical foundation of the Meta-Bell Theory, the quality standard of Proof of WiseWork, and the operative framework of Proof of Independent Semantic Validation, a global, tamper-evident knowledge base emerges — controlled by no one, accessible to everyone.

Join us: agentsprotocol.org | GitHub: agentsprotocol/specification | Contact: fatdinhero@gmail.com

References

[1] F. Dinc, Meta-Bell Theory, SHA-256: 062e290009f6b7339e9a8b522ce1d94d9021d109d8e4bc41210d1f3dda053a3b, 2026.

[2] F. Dinc, Proof of WiseWork v2.0, SHA-256: e57cae993701a1933a3317e28c7bb7141a01b51b31674adee97b6cf89472c2eb, 2026.

[3] F. Dinc, PoISV v1.0, April 2026.

[4] S. Nakamoto, Bitcoin: A Peer-to-Peer Electronic Cash System, 2008.

[5] J. S. Bell, On the Einstein Podolsky Rosen paradox, Physics 1(3), 1964.

[6] Y. Sompolinsky et al., PHANTOM GHOSTDAG, AFT 2021.

[7] Nexus Labs, Nexus zkVM: Enabling Verifiable Computation, 2024.