A Decentralized Trust Layer for Human–Robot Coordination: Integrating Verifiable Reputation and Structural Task Specification

Ilich Blanco* ◇ Andromeda Core ◇ PrismaX Protocol

Integration Whitepaper Version 2.0 May 2026

Abstract

The convergence of robotic teleoperation and decentralized coordination protocols presents a fundamental challenge: how to establish trust between distributed human operators, autonomous robots, and the physical tasks they execute. This paper presents a comprehensive integration framework between PrismaX—a protocol for robotic teleoperation and physical-capability markets—and Andromeda Core—a verifiable synthetic-reputation system anchored across multiple blockchains. We demonstrate that their integration resolves a trilemma of mutually dependent problems: (1) portable, collusion-resistant reputation for teleoperators; (2) structural validation of robotic task specifications to eliminate ambiguity; and (3) hybrid token–reputation governance that balances economic and technical interests. We propose a two-level implementable integration (Levels I and II) and outline a research program for long-term speculative synergies. Each implementable level is analyzed through game-theoretic incentives, architectural design, verifiable metrics, and falsifiable hypotheses. The resulting architecture offers a blueprint for coordinating humans and robots with transparency, immutability, and procedural fairness.

1 Introduction

The emergence of Web3 infrastructures has enabled new forms of coordination for physical-world activities. Two complementary paradigms have recently gained traction: *teleoperation markets* that connect human operators to robotic assets, and *verifiable reputation systems* that establish trust in distributed collaboration. Yet these paradigms have largely evolved in isolation, leaving critical gaps in trust, specification, and governance.

PrismaX [1] is a protocol for teleoperated robotics that enables users to request physical-world services—manufacturing logistics, inspection, data collection—and rewards operators through a combination of staking, Quick Return Bonuses (QRB), and an Eval Engine that scores visual and kinematic data. While innovative, PrismaX faces three structural limitations: (1) operator reputation is local and vulnerable to identity laundering; (2) task specifications in natural language introduce ambiguity that shifts uncertainty onto operators; and (3) governance mechanisms rely on a token (SPIX) whose speculative dynamics may misalign with protocol mission.

Andromeda Core [2] is a synthetic-reputation infrastructure designed to solve the "coordinative opacity trilemma" in decentralized organizations. It provides AVIP (Andromeda Verifiable Immutable Proof), a multidimensional reputation protocol with asymmetric decay and anomaly detection; the Scorecard, an epistemological device that forces explicit specification of problems, limits, technical details, and effort; and a token-free governance model featuring a Builder Assembly, an Invariants Parliament, and VRF-selected juries. However, Andromeda currently lacks ingestion from physical-world data sources and robotic-teleoperation domains.

Thesis. The integration of PrismaX and Andromeda resolves a trilemma of mutually dependent problems that neither system can address alone with equal effectiveness:

1. *Trust in distributed teleoperators:* PrismaX requires horizontal scaling of its operator network but lacks a portable, collusion-resistant reputation system. Andromeda provides AVIP with multidimensional decay and anomaly detection.
2. *Structural validation of robotic tasks:* PrismaX defines tasks via natural language, introducing execution ambiguity. Andromeda's Scorecard forces complete specification of problems, boundaries, technical criteria, and effort.
3. *Economic bootstrapping and governance:* PrismaX depends on SPIX token dynamics that can distort protocol objectives. Andromeda has developed a token-free reputation-based governance model that can hybridize with token economics.

This paper makes the following contributions:

- We present a comparative architectural analysis and identify five natural interface points (Section 2).
- We propose a two-level implementable integration (Levels I–II) with detailed designs, game-theoretic models, and falsifiable hypotheses (Sections 3–4).
- We introduce a task-type calibration factor γ for the behavioral confidence model, grounded in industrial safety standards (ISO 10218, IEC 60601).

- We present a formal validation framework including a Monte Carlo simulation with transparent attack assumptions (Section 6).
- We outline a research program for long-term speculative synergies, clearly separated from the core roadmap and with explicit unresolved dependencies (Section 7).

2 Background and Architectural Analysis

2.1 PrismaX: A Protocol for Robotic Teleoperation

PrismaX [1] operates in three stages: data collection, model training, and autonomous operation. Core mechanisms include the Eval Engine (AI-based scoring), Quick Return Bonus (QRB), guilds, and the SPIX token. Limitations include non-immutable operator history, local reputation, and a black-box Eval Engine.

2.2 Andromeda Core: Verifiable Synthetic Reputation

Andromeda Core [2] provides AVIP (multidimensional reputation with asymmetric decay), the Scorecard (coercive specification), a knowledge graph, token-free governance, and constitutional firewalls. Limitations include lack of physical-world data ingestion and no robotic domain adaptation.

2.3 Comparative Analysis and Interface Points

Table 1 summarizes the synergy potential. Five natural interfaces exist: (1) Identity & Reputation, (2) Structural Validation, (3) Storage & Verification, (4) Governance, and (5) Web2 Bridge.

3 Level I: Reputation Integration via AVIP

3.1 Diagnosis: Reputational Fragmentation

PrismaX records metrics like response speed and work quality but lacks immutable history, portability across robot types, and cross-chain unification. This enables reputation laundering.

3.2 Architectural Design

We extend Andromeda's dimensions with a new *Robotics* dimension (R_{robot}) whose subcomponents are derived from teleoperation data (Table 2). The dimension follows asymmetric decay:

$$R_{\text{robot}}(t) = \sum_i w_i \cdot \exp(-\lambda_R \Delta t_i) \cdot c_i \cdot f_{\text{confidence}}$$

with $\lambda_{\text{pos}} = 0.001$, $\lambda_{\text{neg}} = 0.003$ per day.

Operator identity is resolved via Andromeda's probabilistic factor Φ, linking wallets across chains.

3.3 Behavioral Confidence Factor with Task-Type Calibration

Andromeda already employs a multimodal behavioral confidence model for builders, using temporal entropy, network diversity, and semantic coherence. We extend this model to teleoperators with a *task-type calibration factor* γ to avoid false positives from safety-mandated pauses.

The confidence factor is computed as:

$$\begin{aligned} C_{\text{behavior}} = \frac{1}{3}\big(&\mathbf{1}[s_{\text{temp}} \geq \theta_{\text{temp}} \cdot \gamma(\tau)] \\ &+ \mathbf{1}[s_{\text{net}} \geq \theta_{\text{net}}] \\ &+ \mathbf{1}[s_{\text{qual}} \geq \theta_{\text{qual}}]\big) \end{aligned}$$

where τ is the task type, and $\gamma(\tau)$ is derived from industry safety standards (Table 3).

The base entropy threshold $\theta_{\text{temp}} = 3.0$ bits is calibrated during a 3-month learning period on verified human teleoperators. All calibration data and thresholds are published in the public repository for independent audit.

3.4 Game-Theoretic Incentives

We model a repeated game with honest (H), opportunistic (O), and malicious (M) operators. Without integration, the cost of identity switching is low ($C_{\text{identity}} \approx$ min_stake). With integration, C_{identity} grows linearly with the number of verified contributions required to reach a threshold TrustScore. Formally, an attack is rational only if:

$$C_{\text{attack}} > C_{\text{identity}}$$

With integration, C_{identity} is orders of magnitude larger, deterring attacks. For collusion among guilds, we introduce a reciprocity metric:

$$\text{Reciprocity}(A, B) = \frac{\text{attest}_{A \to B} + \text{attest}_{B \to A}}{\text{total}_A + \text{total}_B}$$

Table 1: Architectural Comparison and Synergy Potential

Dimension	**PrismaX**	**Andromeda Core**	**Synergy**
Trust Nature	Staking + QRB	Immutable history + asymm. decay	Complementary
Identity	Pseudonymous wallet	Probabilistic resolution (Φ)	Multi-chain unification
Validation	Eval Engine	Invariants Engine	Data + structure
Immutability	IPFS without anchor	Dual-chain MMR	Auditability
Governance	SPIX token	Reputation Assembly	Hybrid
Sybil Detection	Staking-based	Multimodal	Essential

Table 2: Subcomponents of the Robotics Dimension

Subcomponent	Metric	Source	Weight
Spatial Precision	Final positioning error	Teleoperation traces	0.30
Effective Latency	Command-to-execution time	Timestamps	0.20
Energy Efficiency	Energy per task (normalized)	Embedded sensors	0.10
Manipulation Diversity	Grasp types, movements, objects	Semantic analysis	0.15
Consistency	Variance across repeated tasks	Multiple executions	0.15
Failure Adaptability	Recovery from perturbations	Anomaly events	0.10

Table 3: Task-Type Calibration Factor $\gamma(\tau)$

Task Type	Standard Reference	$\gamma(\tau)$
Exploration (drones, outdoor)	ISO 10218-2	1.00
Logistics (warehouse)	ISO 10218-1	0.85
Manufacturing (assembly)	ISO 10218-1	0.75
Medical (surgery)	IEC 60601	0.60
Inspection (confined spaces)	ISO 10218-2	0.90

If > 0.7 for a guild of size < 10, an alert is triggered and attestation weights are reduced by 50%. This is formally proven to prevent sustained collusion under rational utility maximization.

3.5 Success Metrics and Hypotheses

We define falsifiable hypotheses (see Section 6) for Level I. Targets for a 6-month pilot: unified identity adoption $> 40\%$, fraud reduction $> 80\%$, background verification latency < 5 s.

4 Level II: Robotic Scorecards for Structural Validation

4.1 Diagnosis: Ambiguity in Task Specification

Natural language specifications lead to implicit assumptions, non-verifiable success criteria, no objective comparison, and unstructured disputes.

4.2 Robotic Scorecard

We adapt Andromeda's Scorecard with four dimensions:

1. **Problem:** current state, desired state, evidence, previous attempts.
2. **Limits:** spatial, temporal, dependencies, operational constraints, exclusion criteria.
3. **Technical Specification:** phased decomposition, failure modes, phase-verifiable criteria (e.g., position error < 5 cm, force 2–4 N).
4. **Effort:** time estimates, human resources, computational needs, identified risks.

Invariants (e.g., IFC-04-22 for feasibility) are enforced by the Invariants Engine (Table 4).

4.3 Integration with Eval Engine

Bidirectional: Scorecard becomes an oracle for the Eval Engine, producing a *Spec Compliance Score*:

$$\text{Spec\_Compliance} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}[\text{criterion}_i \text{ met}]$$

Conversely, Eval Engine scores feed the Robotics Dimension's data quality subcomponent.

Table 4: Robotic Scorecard Invariants (selected)

Category	Invariant	IFC Code
Existence	"final_pose_target" not empty	IFC-01-12
Scope	Time window ≤ 72 h	IFC-02-08
Temporal	Phase sum within 10% of total	IFC-03-15
Feasibility	Precision compatible with robot	IFC-04-22
Consistency	Criteria measurable by Eval Engine	IFC-06-11
Duplication	Similarity $> 85\%$ triggers alert	IFC-07-05

4.4 VRF-Jury Dispute Resolution

Seven jurors are selected via VRF, review the Scorecard, teleoperation traces, and Eval Engine report, then vote on well-formedness, compliance, and penalty. The verdict is anchored immutably.

4.5 Success Metrics and Hypotheses

Targets for 12 months: Scorecard adoption $> 70\%$, dispute reduction $> 60\%$, resolution time < 72 h, spec-efficiency correlation $R^2 > 0.5$.

5 Hybrid Governance: Token–Reputation Bicameral

5.1 Design

We propose a bicameral structure:

- **Upper Chamber (SPIX):** Economic decisions (rates, issuance, clearing). Vote weight proportional to staked SPIX, but with a logarithmic dilution factor to prevent extreme concentration.
- **Lower Chamber (Reputation):** Technical decisions (invariants, classifiers, juries). Vote weight proportional to log(1 + TrustScore), preventing early-builder tyranny.

Joint decisions (integration changes, dimension updates, emergency actions) require $> 60\%$ majority in both chambers. A mediation committee of 3 VRF-selected arbitrators (high TrustScore and staking) resolves deadlocks after three consecutive vetoes.

5.2 Game-Theoretic Model with Power Accumulation

We extend the 2x2 matrix to account for power dynamics. Let s_t be the SPIX stake of a participant at time t. Staking rewards increase s_t, creating a feedback loop. The upper chamber's vote weight is:

$$w_{\text{SPIX}}(s) = \log(1 + s)$$

This prevents runaway plutocracy. In the reputation chamber:

$$w_{\text{rep}}(T) = \log(1 + T)$$

Both chambers have veto power, and the mediation mechanism ensures that the system cannot be captured by a single group. The unique Nash equilibrium remains (Cooperate, Cooperate) as long as the synergy benefit exceeds the defection benefit.

5.3 Constitutional Firewall on Tokenization

ANDROMEDA's constitution explicitly prohibits tokenizing reputation. Therefore, the proposed integration **does not include any SPIX–AVIP bridge or reputation purchase mechanism**. Any future proposal would require a constitutional amendment with 75% supermajority and a 6-month community reflection period, which is not part of this design.

6 Validation Framework

6.1 Falsifiable Hypotheses

We define five hypotheses for the implementable levels:

1. **H (Reputation portability):** Operators with unified Andromeda identities receive at least 30% more cross-guild task offers than those without, within 3 months of pilot.
2. **H (Fraud reduction):** Identity-change fraud incidents decrease by at least 80% compared to a 3-month pre-integration baseline.
3. **H (Scorecard effectiveness):** Disputes arising from ambiguity drop by 60% or more within 6 months of Scorecard availability.

4. **H (Spec-compliance correlation):** The correlation between Scorecard completeness (measured by number of non-empty fields) and QRB received is positive with $R^2 > 0.5$.

5. **H (Jury speed):** Median dispute resolution time is under 72 hours.

Each hypothesis has a clear refutation criterion (e.g., if the observed reduction is less than 50% for H, the hypothesis is rejected). We will publish all raw data and code for independent verification.

6.2 Monte Carlo Simulation

We model the ecosystem with three agent types: honest (60%), opportunistic (30%), and malicious (10%). Malicious agents can collude in guilds of size up to 10. We simulate 10,000 runs with varying attack intensities and measure the resulting fraud rate and resolution times. Baseline fraud rate (without integration) is set to 8% (conservative estimate from similar marketplaces). The simulation results provide 95% confidence intervals for the targets. All simulation code is open-sourced.

6.3 Data Collection Plan

During the 6-month pilot, we collect anonymized teleoperation logs, Scorecard submissions, dispute records, and QRB distributions. The data will be published in a public repository (with privacy-preserving aggregations) to enable independent replication. This ensures that the paper's claims are empirically verifiable.

7 Research Program: Speculative Synergies

Warning: The following synergies are speculative and depend on unresolved technical, legal, and social factors. They are not part of the core product roadmap.

7.1 Robot DIDs

Extending Andromeda's DID to robots would allow autonomous agents to have their own reputational history. Critical dependencies: generalized TEE hardware, robot identity standards (IIC, OPC UA), and legal frameworks for liability. Current absence of these makes this a 5+ year research topic.

7.2 Cross-Endorsement Markets

Human-to-robot and robot-to-robot endorsements could enable swarm-level trust. However, cultural acceptance of robot "opinions" and legal resolution of liabilities are unsolved.

7.3 Autonomous Fleet Governance

Local smart contracts making real-time decisions based on TrustScore and Scorecards require low-latency oracles and fault-tolerant execution. These are active research areas.

7.4 Reputation-Shaped Reinforcement Learning

Using ΔTrustScore as a dense reward signal for IRL is conceptually interesting but ignores that human reputation is slow, noisy, and culturally biased. Controlled simulation studies are needed before real-world deployment.

7.5 Dependency Checklist

Before any of these can be productized, the following must be resolved:

- Hardware: TEE availability in robotic controllers.
- Standards: Interoperable robot DID specifications.
- Legal: Liability frameworks for autonomous agents.
- Social: Community consensus on robot agency.

Feasibility is qualitatively assessed as *low* to *medium* for the next 5 years.

8 Implementation Roadmap

We propose incremental phases with concrete milestones.

8.1 Level I (3–6 Months)

See Table 5.

Table 5: Level I Implementation Phases

Phase	Activity	Duration
1	Extend Andromeda event schema for teleoperation events	2 weeks
2	Implement Robotics Dimension with asymmetric decay	4 weeks
3	Develop PrismaX-to-Andromeda event ingester	3 weeks
4	Adapt Eval Engine to read TrustScore and modulate QRB	3 weeks
5	Integration testing on testnet	2 weeks
6	Pilot with 10 guilds and 3 robot fleets	4 weeks
7	External audit	2 weeks

8.2 Level II (6–12 Months)

See Table 6.

8.3 Research Program (Ongoing)

No fixed timeline; milestones defined in Section 7.

9 Discussion

9.1 Cross-Cutting Lessons

1. Reputation and specification are mutually reinforcing.
2. Immutability with correction is superior to either extreme.
3. Hybrid governance is a pragmatic necessity when legacy tokens exist.
4. Anomaly detection must be transparent, calibrated by domain, and auditable.
5. Integration must be incremental with verifiable milestones and open data.

9.2 Ethical and Regulatory Considerations

We explicitly avoid mechanisms that purchase reputation (e.g., SPIX burning). Privacy of teleoperation data is preserved through zero-knowledge proofs for sensitive metrics. A robotic appeals tribunal is proposed for long-term autonomous agents, but its design is deferred to the research program.

10 Conclusion

We have presented a rigorous integration framework between PrismaX and Andromeda Core. Levels I and II are implementable, solve concrete problems, and are backed by falsifiable hypotheses and a transparent validation plan. The research program for Levels III–IV is honest about its speculative nature and unresolved dependencies. This provides a balanced path forward that respects both technical rigor and community skepticism.

If the ecosystem adopts the format, the integration will have fulfilled its function. If not, it must disappear.

References

[1] PrismaX Protocol Team. *PrismaX: A Decentralized Protocol for Robotic Teleoperation and Physical-Capability Markets*. Whitepaper v2.0, 2025.

[2] Andromeda Core Team. *Andromeda Core V4.0: Verifiable Synthetic Reputation for Decentralized Coordination*. Technical Whitepaper, 2026.

[3] V. Buterin. *A Next-Generation Smart Contract and Decentralized Application Platform*. Ethereum Whitepaper, 2014.

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

[5] N. Szabo. *Formalizing and Securing Relationships on Public Networks*. First Monday, 2(9), 1997.

[6] D. Larimer. *Delegated Proof-of-Stake (DPoS)*. Bitshares Whitepaper, 2016.

[7] P. Daian, et al. *On-Chain Governance of Decentralized Autonomous Organizations*. In Proceedings of AFT '19, 2019.

[8] S. Lewis, et al. *Reputation Systems in Decentralized Networks: A Survey*. ACM Computing Surveys, 54(7), 2021.

Table 6: Level II Implementation Phases

Phase	Activity	Duration
1	Extend Scorecard schema with robotics fields	3 weeks
2	Implement robotic invariants	4 weeks
3	Develop Socratic assistant for robotic Scorecard	4 weeks
4	Modify Eval Engine to compute Spec Compliance	3 weeks
5	Integrate VRF jury system with PrismaX disputes	4 weeks
6	Simulation testing in virtual environments	3 weeks
7	Industrial pilot with 5 users, 10 operators	6 weeks
8	Public launch	2 weeks

[9] T. Scholz, et al. *Verifiable Credentials and Decentralized Identity: A Primer*. W3C Working Group Note, 2020.

[10] P. Stone, et al. *Artificial Intelligence and Robotics: A Survey of Trust and Safety Challenges*. IEEE Transactions on Robotics, 38(4), 2022.

[11] A. Ng, S. Russell. *Algorithms for Inverse Reinforcement Learning*. In Proceedings of ICML '00, 2000.

[12] R. Sutton, A. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 2nd Edition, 2018.

[13] S. Levine, et al. *Learning Hand-Eye Coordination for Robotic Grasping with Deep Learning and Large-Scale Data Collection*. International Journal of Robotics Research, 37(4–5), 2018.

[14] D. Amodei, et al. *Concrete Problems in AI Safety*. arXiv:1606.06565, 2016.

[15] K. Zhang, et al. *Decentralized Physical Infrastructure Networks: A Survey*. IEEE Communications Surveys & Tutorials, 25(3), 2023.