Know Your Agent

Why Trust Infrastructure Replaces Brand in the Agent Economy

Version 3.0 — March 2026

MolTrust / CryptoKRI GmbH, Zürich

Companion paper: MolTrust Protocol Whitepaper v0.4

Executive Summary. The agent economy is real and accelerating. By 2027, automated traffic will exceed human internet traffic. AI crawlers already consume web content at ratios of up to 30,000 times more than they return as human visitors. In financial services, non-human identities already outnumber human employees 96 to one — and across enterprise environments, the ratio reaches 144:1 (Entro Labs, H1 2025). In this environment, brands face a fundamental transformation. AI agents do not respond to logos or reputation narratives. They evaluate verifiable data: authorization records, behavior history, credential provenance. This paper describes why agent trust infrastructure is becoming as foundational as KYC was for human financial transactions — and why it must be built differently: lighter, open, and based on economic necessity rather than regulatory compulsion. The KYA framework aligns with the Singapore IMDA Model AI Governance Framework for Agentic AI (January 2026). MolTrust is building it.

References [1] Prince, SXSW 2026 [6] Neville, a16z crypto Jan 2026

Platform api.moltrust.ch · v1.2.0 · Base L2

— 01 —

Brands in the Agent Economy: Transformation, Not Disappearance

Brands have served a consistent function for over a century: they allow humans to make fast, low-effort quality assessments. A recognized brand signals reliability, accountability, and predictable behavior — without requiring the buyer to verify each transaction from scratch.

This function does not disappear in the agent economy. It transforms.

AI agents do not process brand narratives. They do not respond to visual identity, heritage, or customer loyalty programs. What they evaluate is structured, verifiable data: who authorized this counterparty, what are its operating parameters, what is its track record of compliance?

The question for businesses and platforms is not whether brands remain relevant. It is how the properties that brands have always represented — quality, reliability, authorization, behavioral consistency — can be expressed in a form that machines can verify.

MolTrust is the technical equivalent: a system that translates brand properties into cryptographic credentials, readable and verifiable by any agent, on any platform, without requiring trust in a central authority.

— 02 —

The Scale of the Shift

The transition to agent-mediated interaction is already underway. These are not projections — they are present conditions, documented by Cloudflare's network data covering over 20% of global web traffic.

Metric	Description	Source
2027	Bot traffic exceeds human internet traffic	Prince, SXSW 2026 [1]
10×	Harder to get same Google traffic vs. a decade ago	Cloudflare Blog, July 2025 [2]
750×	Harder via OpenAI vs. old Google	Cloudflare Blog, July 2025 [2]
30,000×	Harder via Anthropic vs. old Google	Cloudflare Blog, July 2025 [2]
100,000:1	Peak crawl-to-refer ratio, leading AI platforms	Cloudflare Year in Review 2025 [4]
96:1	Non-human vs. human identities in financial services	Neville, a16z crypto 2026 [6]
144:1	NHIs outnumber human identities in enterprise (up 44% in 6 months)	Entro Labs, H1 2025 [A]
50%	Organizations that experienced NHI compromise leading to successful cyberattack	Oasis/ESG, 2024 [B]

These ratios reflect a structural shift: AI systems consume content at scale to serve users, without directing those users back to the source. The economics of the web — built on a 30-year-old exchange of content for traffic — have already changed. Agent-mediated commerce will produce the same shift in transactional infrastructure.

CyberArk's 2025 survey of 1,200 security leaders found that 68% of organizations lack identity security controls for AI specifically, and only 23% treat machine identity security as a distinct program. [C] The infrastructure gap is structural and present.

— 03 —

KYA Is Not KYC — It Is Its Natural Evolution

Know Your Customer (KYC) emerged as a response to a genuine problem: financial systems needed a reliable mechanism to verify human identity before allowing participation in regulated markets. The logic was sound — verify before you transact, and maintain an auditable record.

That logic applies equally to AI agents. The agent economy requires an equivalent: Know Your Agent.

KYA draws directly on KYC's core insight. Before an agent transacts, books, trades, or accesses services on behalf of a human principal, the counterparty should be able to verify: who created this agent, who authorized it, what limits apply, and how has it behaved in past interactions?

The implementation, however, must reflect the realities of agent-to-agent interaction: machine speed, global scale, and cost structures that make traditional compliance workflows economically unviable at the transaction level. KYA is KYC adapted for the agent economy — lighter by necessity, open by design, and driven by economic utility rather than regulatory mandate.

> *"The critical missing primitive here is KYA: Know Your Agent. Just as humans need credit scores to get loans, agents will need cryptographically signed credentials to transact — linking the agent to its principal, its constraints, and its liability. Until this exists, merchants will keep blocking agents at the firewall."*
>
> — Sean Neville, co-founder of Circle and architect of USDC; CEO of Catena Labs (a16z crypto, January 7, 2026) [6]

— 04 —

What KYA Requires — Four Pillars

Effective agent trust infrastructure rests on four verifiable properties. These correspond precisely to what brands have always communicated to human buyers — translated into a form that machines can process.

Pillar	Question	MolTrust Implementation
Identity	Who created this agent? Is it the same agent that performed previous transactions?	W3C DID (did:moltrust:...), Ed25519 key pair, Base L2 anchor
Authorization	Whose instructions is this agent executing? What is it permitted to do?	Agent Authorization Envelope (AAE): MANDATE / CONSTRAINTS
Behavior History	Has this agent operated within its stated parameters? Is it trustworthy?	Skill interactions, credential, interaction proofs, trust score (Swarm
Portability	Can any counterparty verify all of the above independently?	Verifiable independent: Ed25519 + JSON-LD, no MolTrust query requ

— 05 —

Where KYA Becomes Critical: Selected Industries

The following represents an initial set of industries where agent trust infrastructure is already operationally relevant. Every sector that has digitized decision-making will eventually require equivalent infrastructure. The question is timing and who establishes the standard.

First Wave — Operationally Relevant Today

Commerce

Autonomous shopping agents negotiate and complete purchases on behalf of human principals. Merchants have no mechanism to verify whether an agent is authorized, what payment limits apply, or whether it has a history of disputed transactions. Pre-transaction verification of agent credentials — including spend limits, authorized categories, and principal identity — is the minimum viable trust layer.

Financial Services

Trading agents and portfolio management systems execute transactions at speeds and volumes beyond human oversight. Delegation chains — where a human principal authorizes an agent, which delegates to a sub-agent — create liability gaps that existing identity infrastructure cannot resolve. The Agent Authorization Envelope (AAE) addresses this by encoding approval thresholds and step-up requirements directly in the credential.

Prediction Markets and Data Services

AI-powered prediction services operate with minimal accountability for claimed track records. Without a timestamped, immutably anchored commitment mechanism, any historical performance claim is unverifiable.

Second Wave — Emerging Relevance

Healthcare

Diagnostic agents, prescription management systems, and clinical decision support tools operate where agent identity and authorization are matters of patient safety and regulatory compliance.

Public Services

Government-facing agents that file documents, access citizen records, or execute administrative processes require pre-transaction verification infrastructure that existing government identity systems are not designed to provide.

Climate and Critical Infrastructure

Energy trading agents, grid management systems, and emissions reporting infrastructure operate at the intersection of commercial and public interest. The consequences of unauthorized or misconfigured agent behavior extend beyond financial loss.

— 05b —

Real-World Deployments: KYA Infrastructure in Production

The following case studies are drawn from MolTrust's live platform (api.moltrust.ch, v1.2.0). They are not hypothetical scenarios. Each represents a deployed credential vertical with active endpoints, passing test suites, and on-chain anchoring on Base L2.

Case Study 1: Prediction Markets — Verifiable Track Record

Problem. AI-powered prediction services operate without accountability for claimed performance. Any agent can assert a 78% accuracy rate. Without a timestamped, immutably anchored commitment mechanism, that claim is unverifiable — and indistinguishable from fabrication.

Implementation. MolTrust's Prediction Markets vertical introduces the PredictionTrackCredential: a W3C Verifiable Credential issued only after a prediction has been cryptographically committed before the event and verified after outcome settlement via independent data sources. The credential encodes the agent's DID, the committed prediction hash, the settlement timestamp, and cumulative accuracy — all anchored on Base L2.

Outcome. A counterparty querying an agent's prediction credential can verify: (1) the commitment existed before the event, (2) the claimed accuracy is derived from on-chain settlement records, not self-reported, (3) the agent's DID is consistent across all transactions.

KYA relevance: Behavior History pillar in production.

Case Study 2: Commerce — Spend Limit Enforcement via AAE

Problem. Merchants have no mechanism to verify whether an agent is authorized, what payment limits apply, or whether the agent has a history of disputed transactions.

Implementation. MolTrust's Shopping vertical issues BuyerAgentCredential credentials embedding a full AAE. The AAE encodes: MANDATE (permitted purchase categories, allowed merchants), CONSTRAINTS (spend ceiling per transaction and per 24h window, geographic restrictions), and VALIDITY (expiry timestamp, revocation endpoint).

Outcome. An agent with a €500/transaction limit attempts a €1,200 purchase. The merchant's verification endpoint reads the AAE constraints and pauses the transaction pending human approval — before any funds move.

KYA relevance: Authorization pillar in production.

Case Study 3: Brand Protection — Delegation Chain Verification

Problem. In agent-mediated commerce, an agent claiming to represent an authorized reseller may be fabricating that authorization. There is currently no standard mechanism to verify multi-hop delegation chains between commercial agents.

Implementation. MolTrust's Salesguard vertical issues ProductProvenanceCredential (issued by the brand to authenticated products) and AuthorizedResellerCredential (issued by the brand to verified distributors). A verifier walks the delegation chain from product to brand in a single API call.

Outcome. The verification result is binary and machine-readable: this agent is — or is not — operating within a brand's authorized distribution network. No human review required.

KYA relevance: Identity + Authorization pillars combined.

— 05c —

The Economics of Trust Failure

KYA adoption will not be driven primarily by regulatory mandate. It will be driven by the cost differential between trust failure and trust infrastructure. The question is not whether organizations will invest in agent trust verification — it is whether they will do so proactively or reactively, after absorbing the cost of failures.

What Trust Failure Costs

Identity-based fraud is not a future risk. Javelin Strategy & Research (2025) found that account-takeover fraud reached $15.6 billion in losses in 2024, up from $12.7 billion in 2023 — a 23% year-on-year increase driven primarily by automated credential exploitation. [D] TransUnion's H2 2025 Global Fraud Report found that businesses lost on average 8% of annual revenues to fraud, with digital account-takeover volume growing 21% from H1 2024 to H1 2025. [E]

The Oasis/ESG study found that nearly 50% of organizations experienced non-human identity compromises that led to successful cyberattacks in 66% of those cases. [B] These figures predate the widespread deployment of autonomous AI agents as transaction initiators.

The Numbers: A Single Use Case

Consider a merchant processing 10,000 agent-initiated transactions per day. The TransUnion baseline of 8% revenue loss to fraud suggests that without a trust verification layer, fraud exposure scales directly with transaction volume.

Cost Item	Per Incident	At 0.1% Fraud Rate — conservativ
Chargeback processing fee	€25–90	€250–900/day
Merchant penalty (repeat threshold)	€500–2,500/month	up to €2,500/month
Lost goods / service cost	variable	variable
Total monthly exposure	—	€10,000–30,000+

KYA Cost Item	Per Verification	At 10,000 verifications/day
On-chain anchor (Base L2)	< €0.01	< €100/day
Credential verification call	< €0.001	< €10/day
Total monthly KYA cost	—	< €3,300/month

The Structural Argument

The underlying economics of cryptographic verification are structurally asymmetric: the marginal cost of verifying a credential approaches zero at scale. A signed JSON object can be validated against a public key in microseconds, with no intermediary, no human review, and no per-verification fee to a central authority.

This is a property of the cryptographic primitives — Ed25519 signatures, W3C Verifiable Credentials, and immutable on-chain anchors — that KYA infrastructure is built on. Any implementation using these standards inherits the same cost structure.

KYA does not eliminate trust — it relocates it. Trust shifts from the counterparty's brand or reputation to the credential issuer and the cryptographic proof chain. The issuer remains a trust anchor. What changes is that this anchor is explicit, auditable, and machine-verifiable rather than implicit and unverifiable at transaction speed.

KYA adoption will not wait for regulatory mandate. The economic pressure is already present. What regulators will eventually require, the market will have already built — because the alternative is operating without a trust layer in an environment where counterparties are increasingly autonomous, fast, and impossible to verify by human review.

— 05d —

The Existing Landscape — and Why It Does Not Solve KYA

Agent trust infrastructure does not emerge into a vacuum. Several categories of existing solutions address adjacent problems. None address the core KYA requirement: pre-transaction,

machine-verifiable, portable credential verification for autonomous agents operating across platforms and jurisdictions.

Traditional Identity and Access Management (IAM)

Enterprise IAM systems (OAuth 2.0, SAML, LDAP-based directories) were designed for human users authenticating to organizational systems. They assume a centralized authority, a human in the authentication loop, and a static organizational boundary. Agent-to-agent commerce requires none of these. IAM systems do not provide portable, verifier-independent credentials that a counterparty can validate without querying the issuing organization's infrastructure.

Web3 Identity and Decentralized Identifier Frameworks

The W3C DID and Verifiable Credentials specifications provide the correct technical foundation — and MolTrust builds directly on them. The gap is application-layer specificity: general-purpose DID infrastructure does not encode agent-specific properties (mandate scope, spend limits, delegation chains, behavior history) in a form that agent commerce workflows can consume.

AI Framework-Native Trust Mechanisms

Major AI agent frameworks implement internal trust mechanisms — permission scopes, tool restrictions, sandboxing. These operate within a single framework's boundary and are enforced by the framework operator. An agent's claimed permissions within Framework A cannot be verified by a counterparty operating on Framework B without a shared trust layer.

The Gap

The common limitation across all three categories is portability under adversarial conditions: the ability for a counterparty who does not know, trust, or share infrastructure with the agent's operator to verify the agent's identity, authorization, and behavior history — in milliseconds, without a phone call, without a human review, and without trusting the agent's own assertions. That is the gap KYA infrastructure fills.

— 05e —

Adoption Barriers: What KYA Must Navigate

KYA infrastructure does not enter a frictionless market. Understanding the barriers to agentic AI adoption is a prerequisite for understanding the adoption curve of the trust layer that enables it.

The Current State of Agent Deployment

Gartner's September 2025 survey found that only 15% of enterprises were actively exploring fully autonomous agents without human oversight. [F] Dynatrace research found that approximately half of AI agent projects remain in proof-of-concept stages, with around 70% of agent outputs still subject to human verification before action. [F] The primary blocker: security, privacy, and compliance concerns, cited by 52% of Dynatrace respondents.

The barrier to agentic AI adoption and the demand for KYA infrastructure share the same root cause: organizations cannot deploy agents at scale without a portable, machine-verifiable trust layer that enforces authorization cryptographically — not by organizational policy alone.

Three Specific Barriers KYA Addresses

Security and unauthorized access

The 52% of organizations citing security as the primary blocker are responding to a real risk: CyberArk found that 68% lack identity security controls for AI agents specifically. [C] KYA addresses this directly: AAE-embedded spend limits and authorization scopes enforce constraints before a transaction executes, not after an incident is detected.

Governance and accountability gaps

Teradata's survey of 500+ business leaders found that data governance and compliance gaps are among the top three barriers to scaling agent deployments. [G] The KYA trust chain — from human principal to agent instance, with every delegation step recorded in a cryptographically signed credential — is a governance architecture, not just a security feature.

Integration complexity

MolTrust's verifier-independent design addresses this at the architectural level: any counterparty can validate a credential without integrating with MolTrust's infrastructure. The W3C VC standard ensures compatibility with any system that speaks JSON-LD and Ed25519.

What KYA Does Not Solve

KYA infrastructure eliminates the trust verification problem. It does not eliminate the underlying complexity of deploying capable agents in production environments — model quality, tool reliability, orchestration architecture, and change management remain the responsibility of the deploying organization. MolTrust provides the trust layer. The agent stack is the customer's.

— 05f —

Technical Integration: Two Deployment Patterns

KYA infrastructure integrates at two distinct points in an agent's execution flow. The following patterns reflect MolTrust's current production implementation.

Pattern 1: Pre-Transaction Authorization Check

An agent presents its credential (including embedded AAE) to a counterparty's verification endpoint before initiating a transaction. The verifier evaluates: (1) credential signature validity against the issuer's public DID key, (2) AAE CONSTRAINTS block — spend ceiling, permitted categories, geographic scope, (3) VALIDITY block — expiry and revocation status, (4) on-chain anchor match. Latency: sub-100ms in

production under normal Base L2 conditions.

```
Agent → POST /travel/verify {credential, proposed_action}
Verifier → checks AAE.CONSTRAINTS against proposed_action
Verifier → resolves DID document, verifies Ed25519 JWS proof
Verifier → queries on-chain anchor (Base L2 read, ~50ms)
Verifier → returns {verified: true, constraints_met: true, expires_at: ...}
```

Pattern 2: Delegation Chain Verification

A sub-agent acting on behalf of a parent agent presents a credential chain. The verifier walks the chain from the sub-agent's credential to the parent agent's credential to the human principal's root authorization — each link cryptographically signed by the upstream issuer's DID key. The sub-agent's AAE encodes only the delegated scope (a subset of the parent's mandate).

```
SubAgent credential  →  signed by ParentAgent DID
ParentAgent credential  →  signed by HumanPrincipal DID
HumanPrincipal DID  →  resolvable on Base L2
Verifier walks chain:  GET /trust/chain/:did
```

Both patterns are live at api.moltrust.ch and testable via the MCP server (48 tools, compatible with any MCP-enabled AI assistant or development environment).

— 06 —

The Regulatory Question

KYC was not adopted voluntarily across financial markets. It was established through successive regulatory frameworks — the Bank Secrecy Act, FATF recommendations, the EU's Anti-Money Laundering Directives — each responding to documented failures in the absence of standardized identity verification.

The EU AI Act introduces obligations around transparency and accountability for AI systems in high-risk domains. MiCA establishes identity requirements for crypto asset service providers. DORA mandates operational resilience for financial entities using third-party technology. None of these frameworks yet address agent-to-agent interaction directly. That gap will close.

Singapore IMDA Model AI Governance Framework Alignment

MolTrust implements the Singapore IMDA Model AI Governance Framework for Agentic AI (January 2026, WEF) — the world's first governance standard for autonomous AI systems. The four MGF dimensions map directly to the KYA pillars:

MGF Dimension	KYA Pillar	MolTrust Implementation
Agent Identity & Integrity	Identity	W3C DID with Ed25519 cryptographic binding; Base L2 tamper-evident ancho
Authorization & Scope	Authorization	AAE: MANDATE / CONSTRAINTS / VALIDITY — machine-readable, cryptogr
Accountability Chain	Behavior History	Five-party trust chain recorded in every credential; AAE approval thresholds
Technical Controls	Portability	W3C VC Ed25519 JWS proof; revocation endpoints; verifier-independent valid

— 06b —

Market Adoption vs. Regulatory Timeline

The Correct Sequence

KYC was not adopted voluntarily — it required successive regulatory mandates. KYA will follow a different path, because the economic incentive structure is different.

KYC solved a regulatory problem: preventing financial crime in human-to-human transactions. The adoption cost was high (human review, institutional intermediaries, multi-week onboarding), and no individual actor had strong economic incentive to implement it unilaterally. KYA solves an operational problem: enabling agent-to-agent commerce at speed and scale. The adoption cost is low (millisecond verification, sub-cent anchoring, self-service registration). Every actor that deploys agents without KYA infrastructure is already paying the cost of that absence.

This means KYA adoption is pull-driven, not push-driven. Platforms will require KYA-compliant agents because it reduces their operational risk. Merchants will prefer credentialed agents because it eliminates chargeback exposure. Financial infrastructure will route preferentially to agents with verified authorization records.

Regulators will arrive after the market has established the standard. The EU AI Act (enforcement August 2026), FATF guidance on autonomous agent identity, and emerging frameworks in APAC jurisdictions will codify what the market has already built — not create it from scratch.

Implications for Positioning

MolTrust is not building compliance infrastructure for a future regulatory requirement. It is building the operational layer for agent commerce that the market needs now. Regulatory alignment is a consequence of building correctly, not a prerequisite for adoption. The competitive advantage for

organizations that move first is operational capability: the ability to deploy agents that counterparties will accept, transact with, and trust — before the majority of the market has built that capability.

— 07 —

MolTrust: KYA Infrastructure, Available Today

MolTrust provides the foundational components of KYA infrastructure, live at api.moltrust.ch. The current implementation (v1.2.0) covers eight credential verticals: Core Identity, Commerce, Travel, Skill Verification, Prediction Markets, Brand Protection, Music, and Sports Integrity.

48 tools are available via the Model Context Protocol (MCP), compatible with any MCP-enabled AI assistant or development environment. All credentials are immutably anchored on Base L2, providing tamper-evident, publicly verifiable transaction records. Agent registration is self-service, operational within minutes, with no human review required.

Every issued credential now embeds an Agent Authorization Envelope (AAE) — a structured, cryptographically bound object that encodes the agent's mandate, constraints, and validity period. The AAE enables pre-transaction verification: before an agent acts, any counterparty can evaluate the envelope against the proposed action and determine — in milliseconds — whether the operation is permitted, requires step-up authentication, or needs human approval.

Phase 1 of the MolTrust Swarm Intelligence Protocol — introducing peer-propagated trust scores through SkillEndorsementCredentials — is LIVE as of March 2026. The technical design is Technical documentation is available in the companion paper: MolTrust Protocol Whitepaper v0.4 (March 2026), moltrust.ch

— 08 —

Contact and Next Steps

Platform integration and Early Access	`api.moltrust.ch/auth/signup`
Partnership and investment inquiries	`info@moltrust.ch`
Technical documentation	`api.moltrust.ch/docs`
Protocol Whitepaper v0.4	`moltrust.ch/MolTrust_Protocol_Whitepaper_v0.4.pdf`
KYA Whitepaper (this document)	`moltrust.ch/MolTrust_KYA_Whitepaper.pdf`

References

[1] Prince, M. (2026). "The Internet After Search." SXSW, March 2026.

[2] Prince, M. (2025). "Content Independence Day." Cloudflare Blog, July 1, 2025.

[4] Cloudflare, Inc. (2025). "The 2025 Cloudflare Radar Year in Review." December 2025.

[5] Cloudflare, Inc. (2025). "The crawl-to-click gap." October 2025.

[6] Neville, S. (2026). "We'll go from KYC to KYA." a16z crypto, January 7, 2026.

[7] W3C Decentralized Identifiers (DIDs) v1.0. w3.org/TR/did-core/

[8] W3C Verifiable Credentials Data Model v2.0. w3.org/TR/vc-data-model-2.0/

[9] x402 Payment Protocol Specification. x402.org

[10] MolTrust Swarm Intelligence Whitepaper v4. March 2026. moltrust.ch/whitepaper.html

[11] IMDA, Singapore. "Model AI Governance Framework for Agentic AI." January 22, 2026.

[12] EU AI Act. Regulation (EU) 2024/1689. Enforcement begins August 2026.

[13] Republic of Korea. "AI Basic Act." Effective January 2026.

[A] Entro Labs. "NHI & Secrets Risk Report — H1 2025." July 28, 2025. Methodology: enterprise-environment telemetry, Jan–Jun 2025.

[B] Oasis Security / ESG. "Managing Non-Human Identities." December 31, 2024. ESG quantitative practitioner survey.

[C] CyberArk. "2025 State of Machine Identity Security Report." June 16, 2025. Survey of 1,200 security leaders, 6 countries (Censuswide).

[D] Javelin Strategy & Research / AARP. "Identity Fraud Report 2024." March 25, 2025. Survey of 5,023 U.S. adults.

[E] TransUnion. "H2 2025 Global Fraud Report." October 6, 2025. Aggregated global transaction and fraud-event data.

[F] AI-CERTS (synthesizing Gartner Sep 2025 and Dynatrace). January 25, 2026.

[G] Teradata. "Survey on AI Agents and Customer Experience." October 7, 2025. Survey of 500+ business leaders.

