consult hyperion
consulting by fime

From KYA to Runtime Trust: Closing the Gap in Agentic Commerce.

No identity, no trust; no trust, no autonomous commerce.

Date:
September, 2026

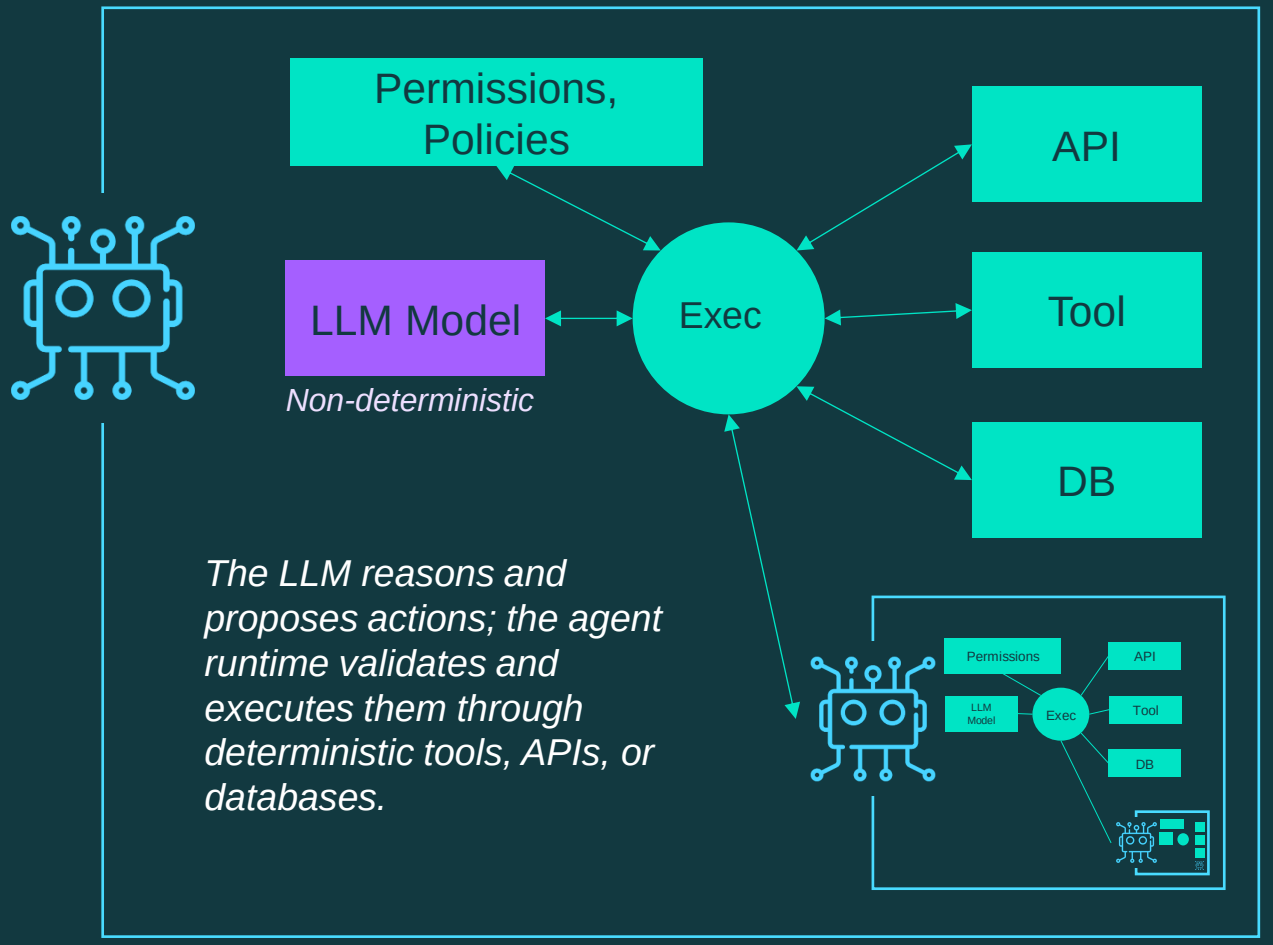
Presentation to:
W3C / GS1

consult hyperion

What is an agent?

An AI agent is an LLM-based system deployed in the cloud, on a device, or on-premises, with instructions and access to tools—such as APIs, database, or Tools (e.g. MCP) —to perceive information, make decisions, and act on a user's behalf.

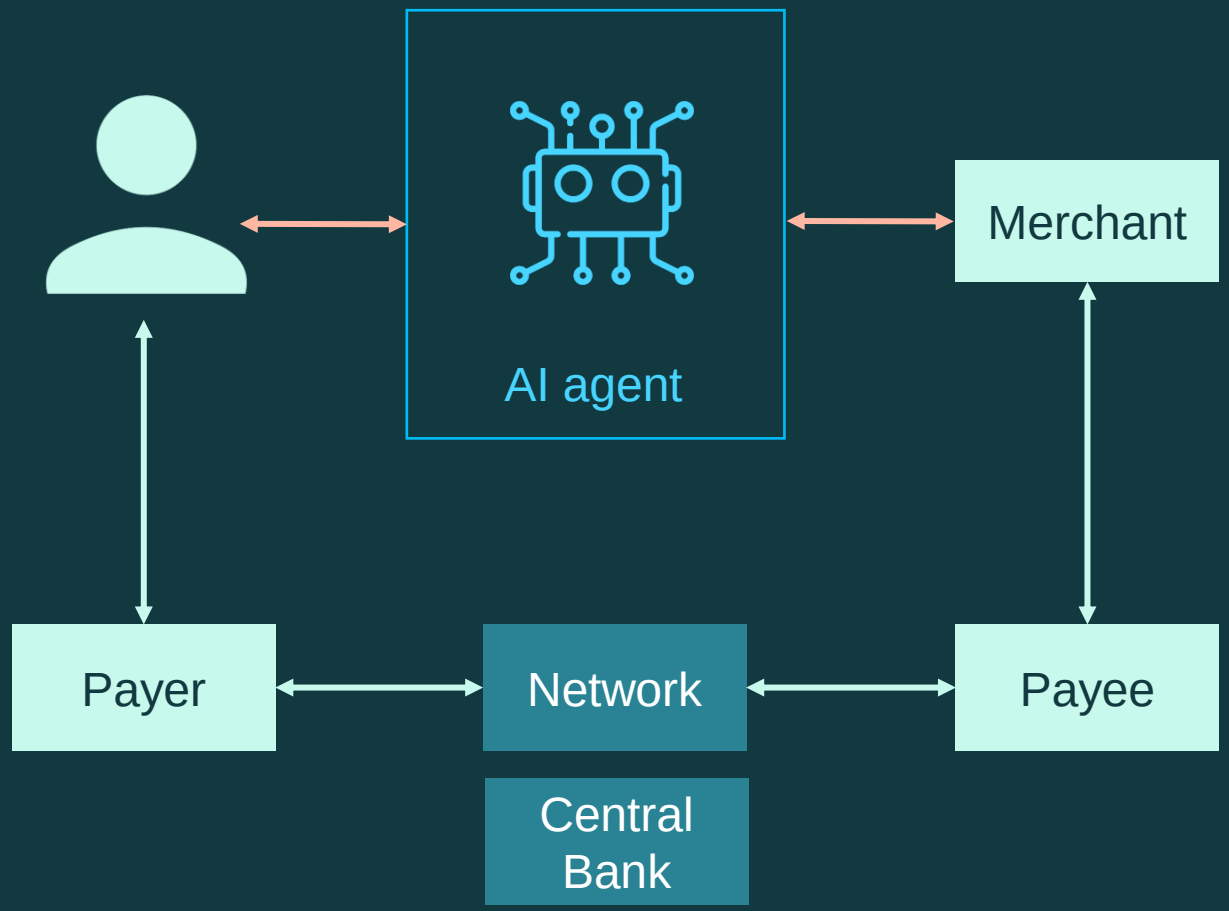
Agents may also delegate tasks to specialized sub-agents, creating nested or multi-agent workflows



Nested structure

consult hyperion

What agentic AI means for the payment ecosystem ?



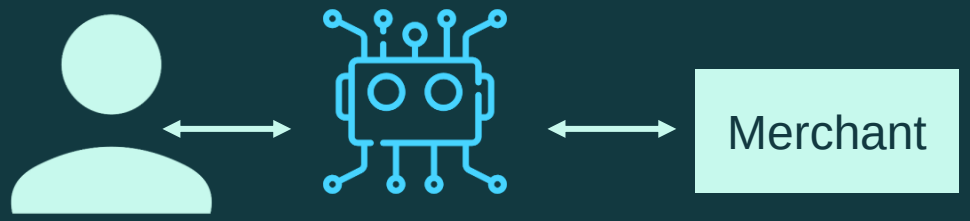
The classic Four-Corner Model in payments involves four key actors—Cardholder, Merchant, Acquirer, and Issuer—who facilitate secure transaction flows and settlement.

Emerging agentic commerce introduces a fifth actor, the autonomous AI Agent, which acts on behalf of the user to initiate and manage transactions.

This new role transforms the traditional model by adding new interactions and responsibilities, requiring updated security, compliance, and operational standards to support AI-driven commerce.

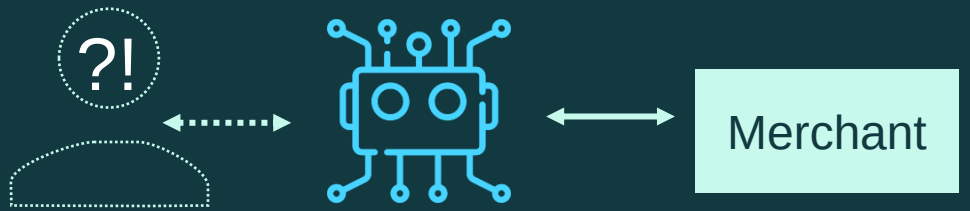
consult hyperion

Overview of the emerging agentic AI commerce ecosystem Paradigm shifting.



1. Immediate: Human-In-The-Loop

- The user actively supervises and authorizes transactions.
- The AI agent suggests options and seeks explicit approval before executing payments.



2. Delegation: When Human is not present

- The user pre-authorizes the AI agent to act autonomously under predefined conditions.
- The agent places orders, manages payments, and completes transactions without real-time human input.

consult hyperion

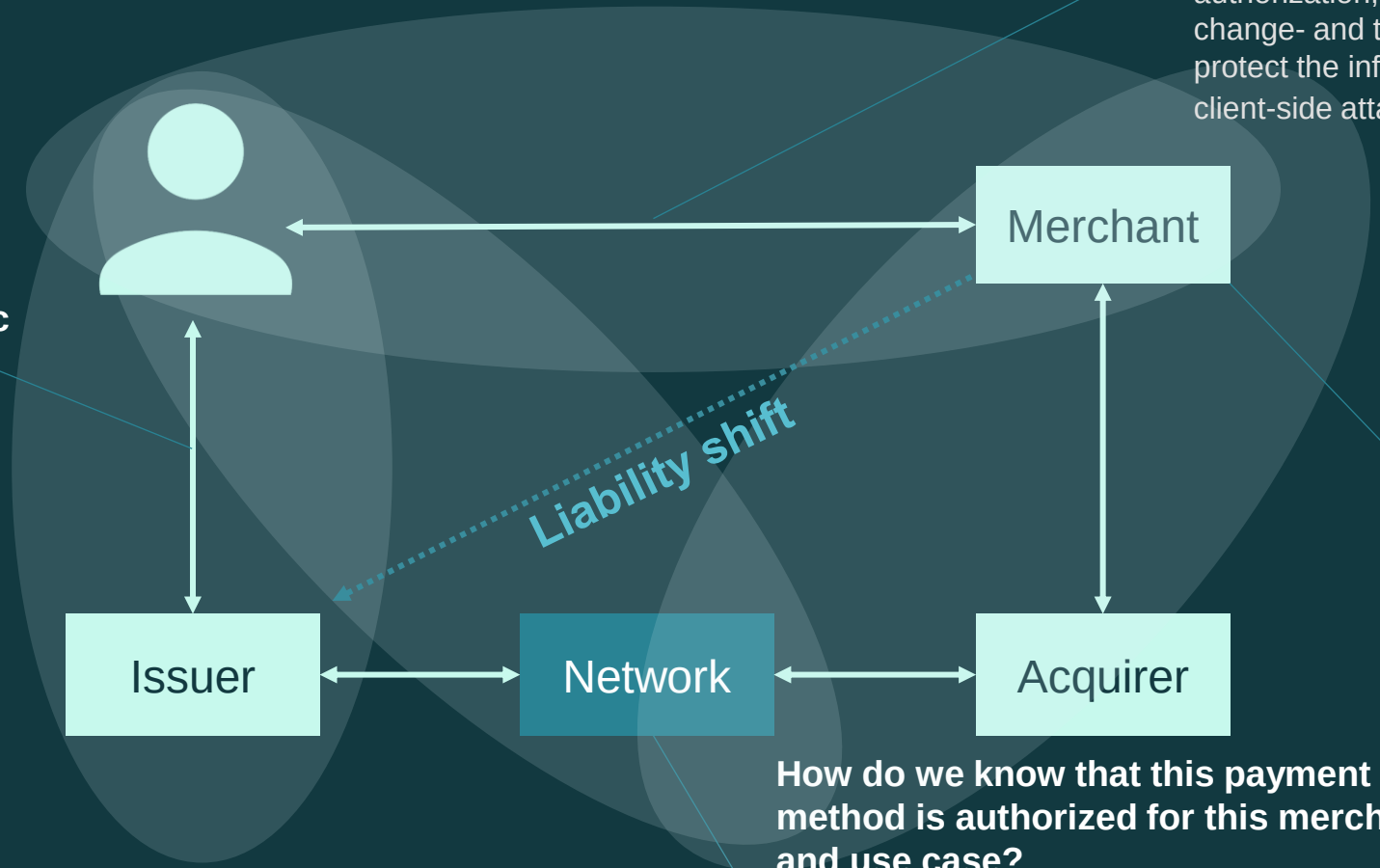
Today's ecommerce trust baseline.

Consent integrity

How do we know that the payment information shown to the user has not been manipulated?
PCI DSS requirements address the security of the payment page and its browser-side scripts. Script authorization, integrity controls, inventories, and change- and tamper-detection mechanisms help protect the information rendered in the browser from client-side attacks.

How do we know that the user approved this specific transaction?
Strong Customer Authentication, EMV 3-D Secure, and passkeys* bind the authentication event to transaction details such as the merchant, amount, payment method and context. This is commonly described as dynamic linking.

User consent



How do we know that the merchant is legitimate and accountable?
Merchant onboarding and due diligence by the acquirer and payment network establish the merchant's identity, business legitimacy, and accountability. Network and acquirer rules provide recourse when fraud, misuse, or non-compliance occurs.

Merchant legitimacy

How do we know that this payment method is authorized for this merchant and use case?
EMV payment tokens replace the underlying card number and are constrained by token-domain controls, such as the merchant, device, channel or transaction type. These controls limit use outside the intended context.

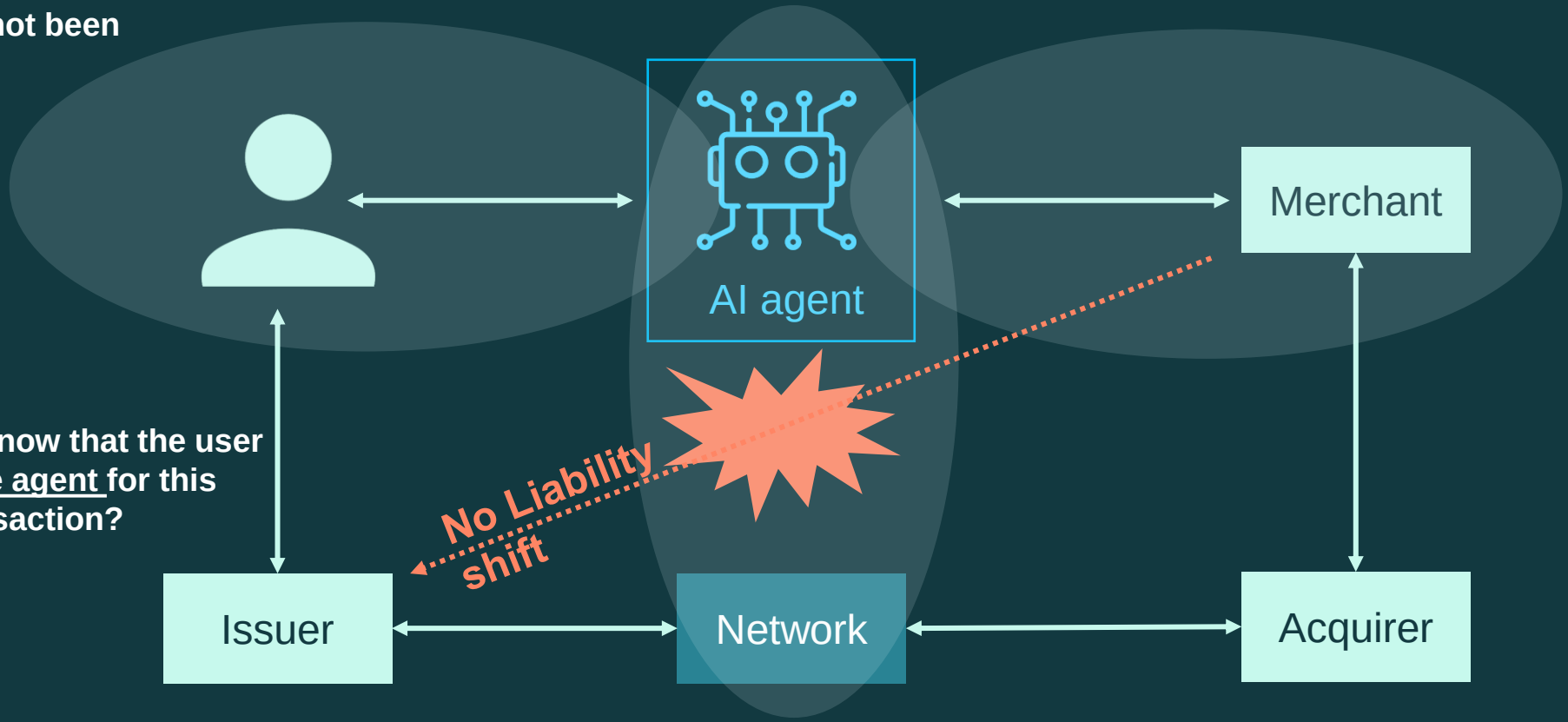
Payment authorization

consult hyperion

Impact of agentic commerce in trust model

How do we know that the payment information shown to the user has not been manipulated?

How do we know that the ~~merchant~~ agent is legitimate and accountable?



How do we know that the user approved the agent for this specific transaction?

How do we know that this payment method is authorized for this ~~merchant~~ agent and use case?

consult hyperion

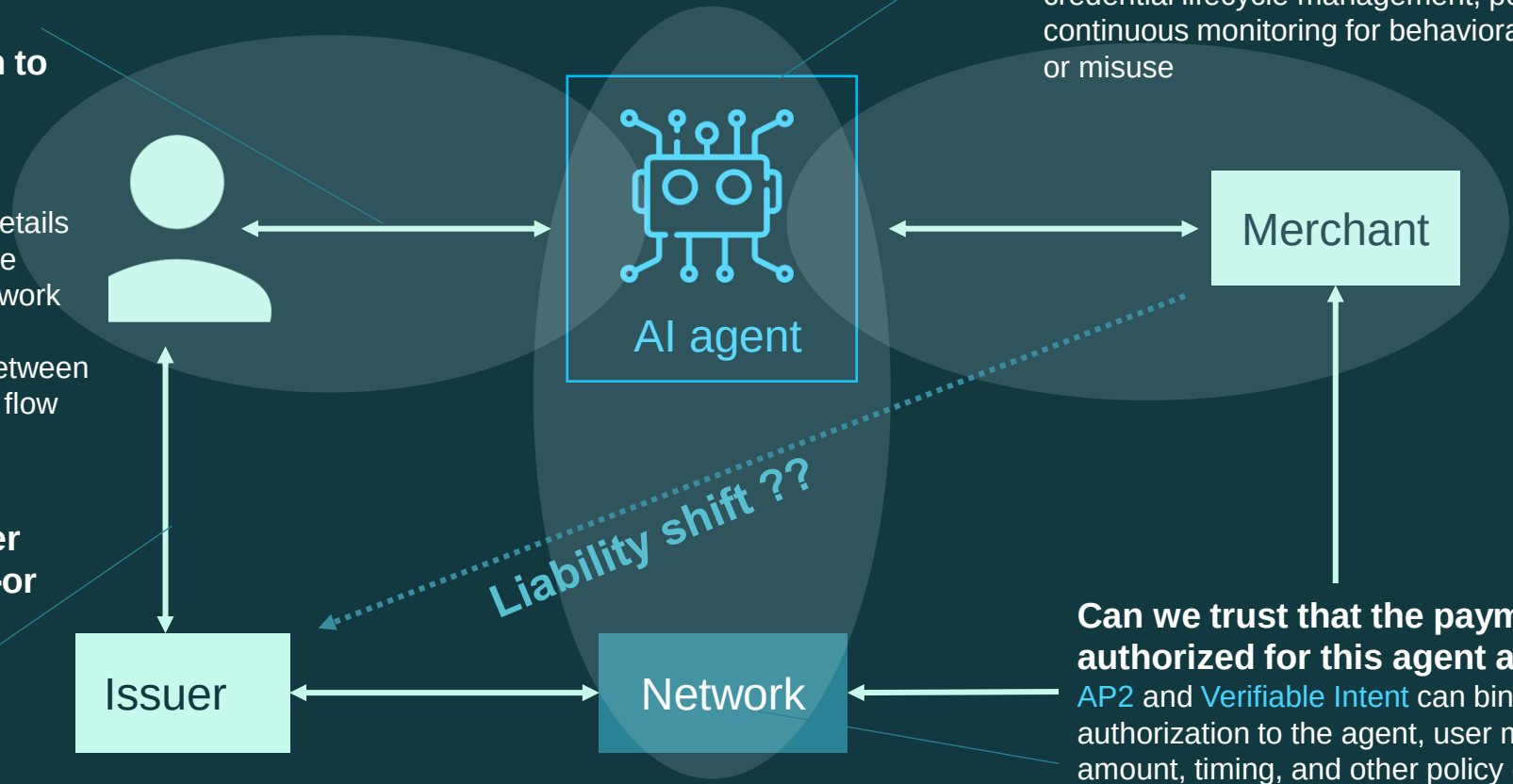
Emerging trust Framework

How do we know which agent is acting, who operates it, and whether it should still be trusted?

KYA establishes an agent's identity, operator, principal, capabilities, and delegated authority during onboarding. Its trust status is then maintained through credential lifecycle management, policy checks, and continuous monitoring for behavioral drift, compromise, or misuse

How do we know that the payment information shown to the user has not been manipulated?

AP2 and Verifiable Intent bind the merchant basket and transaction details into signed, verifiable evidence. The merchant, wallet, and payment network can independently check that the transaction has not been altered between the agent, merchant, and payment flow



How do we know that the user authorized this transaction—or authorized the rules under which it was executed?

AP2 and Verifiable Intent provide cryptographic evidence of the user's mandate, consent, and delegation.

Can we trust that the payment method is authorized for this agent and use case?

AP2 and Verifiable Intent can bind payment authorization to the agent, user mandate, merchant, amount, timing, and other policy constraints. Payment networks and issuers can additionally provision transaction-specific or agent-specific tokens, reducing the value of credentials used outside their intended scope.

consult hyperion

What is Know Your Agent (KYA) ?

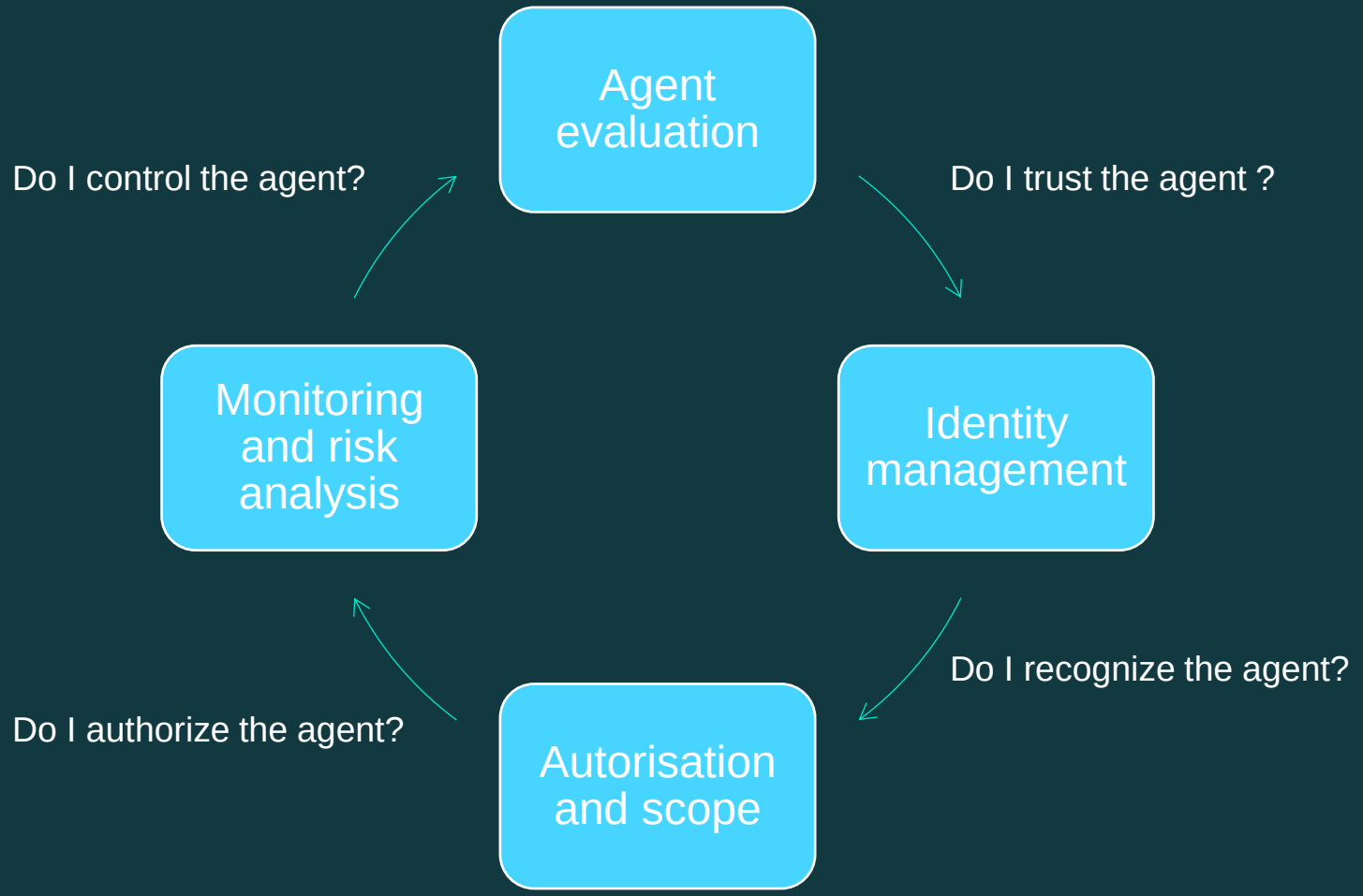
KYA in four steps

1.Evaluate the agent
Assess the agent, its operator, deployment, model, purpose, and controls against applicable regulations, policies, and risk requirements before onboarding.

2.Manage its identity
Assign the agent a unique, verifiable identity and use interoperable standards—such as DIDs, verifiable credentials, or signed tokens—to share and verify it.

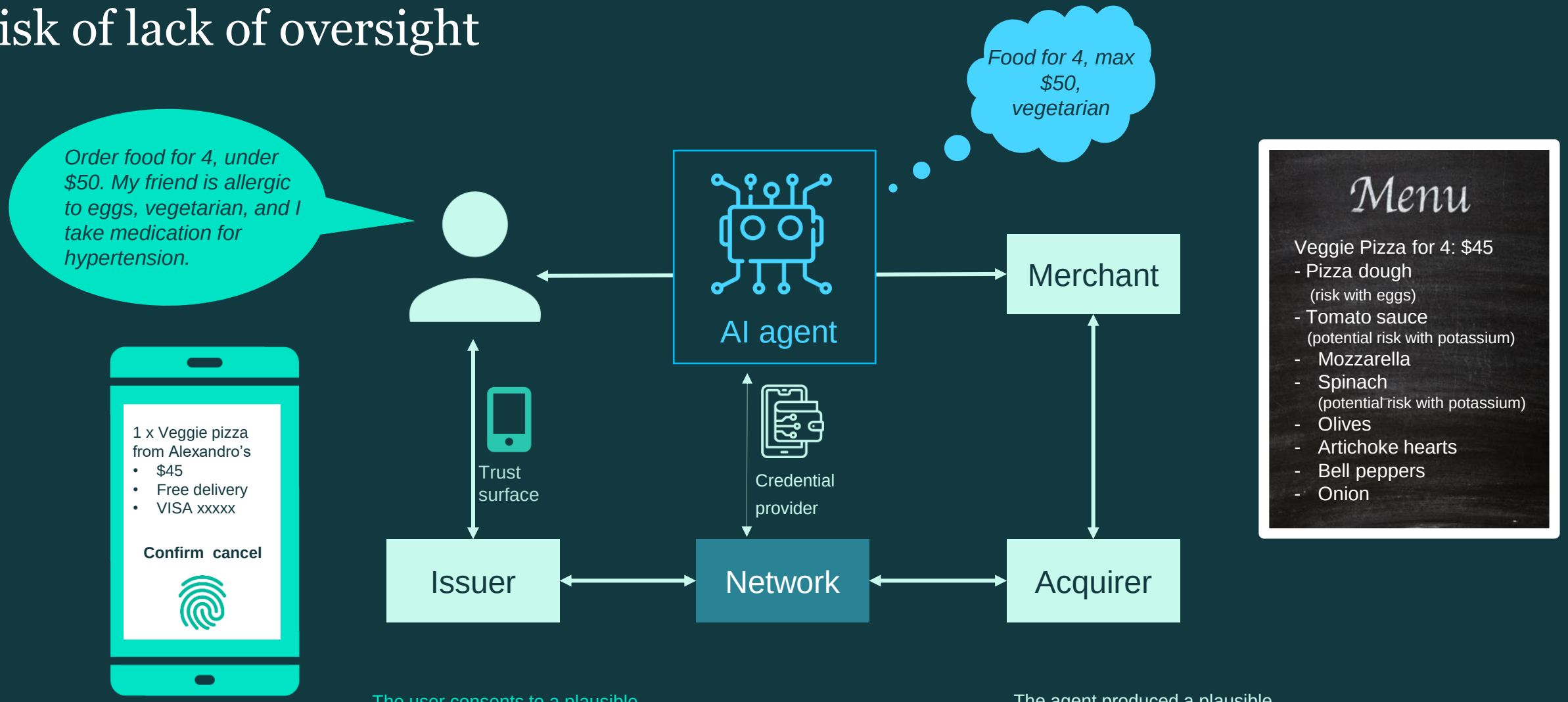
3.Define authorization and scope
Bind the agent to explicit limits: what it may do, on whose behalf, with which tools or payment methods, for which counterparties, and within what amount, time, or geographic boundaries.

4.Monitor behavior and manage risk
Continuously observe the agent's actions, detect anomalies or policy violations, and adapt, suspend, or revoke its authorization when risk changes.



consult hyperion

Risk of lack of oversight



The user consents to a plausible offer, but the consent no longer reflects the full original intent because allergy and health constraints were reduced during agent interpretation.

The agent produced a plausible commercial match and a valid-looking consent request, while silently dropping critical safety or health constraints from the original user intent

consult hyperion

What remains to be covered.

Agent oversight is required to

- **Verify agent identity and monitor behavior** to detect drift from the approved KYA profile and authorized scope.
- **Audit intent interpretation and execution** to ensure cart construction, merchant selection, and checkout decisions comply with user constraints.
- **Assess regulatory compliance** in merchant and payment-method selection, including fairness, contestability, and applicable regional rules.

Intent drift and reduction.
User intent is compressed into model representations, then often simplified to fit tool schemas and privacy minimization rules, which can weaken or drop important constraints.

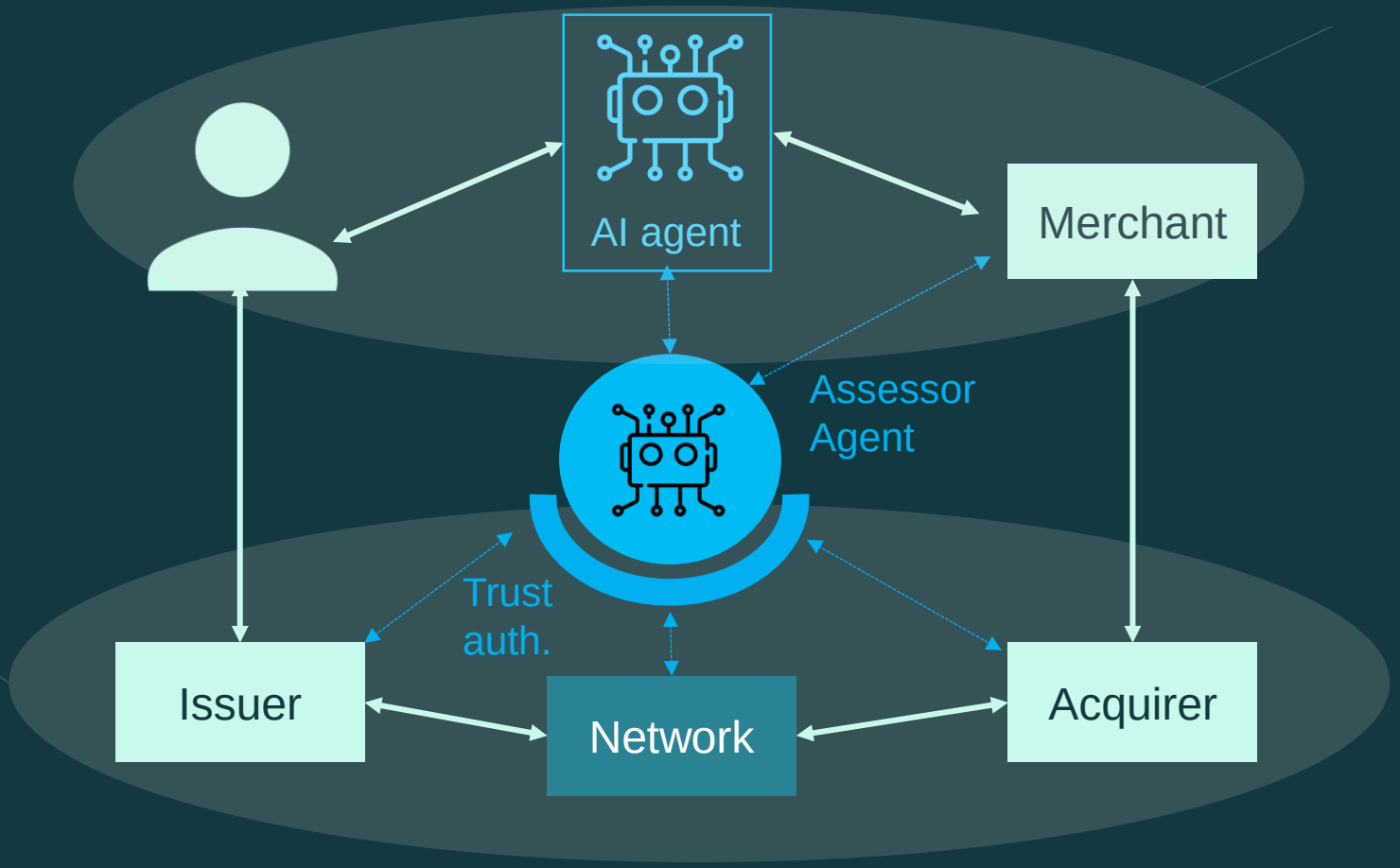
Bias, hallucination, and poisoning.
The agent may hallucinate, inherit bias, or be manipulated through poisoning or prompt injection, leading to actions that no longer faithfully reflect the user's intent or best interests.

Biased or manipulated negotiation.
Agent-to-merchant negotiation may be steered, echoed, or biased, leading to offers, rankings, or prices that no longer serve the user's best interests.

consult hyperion

Additional trust in Agentic AI commerce : FACT and ATLAS

Independent trust layer for continuous verification, oversight, and auditable trust signals



FACT Verify

Independent Trust Assessor check before checkout

- Intent consistency
- Fairness
- Regulatory compliance
- KYA standing, and rulebooks.

Maps to ATLAS direct/synchronous and indirect/asynchronous assessment modes.

FACT Evidence

The audit trail / verifiable record: signed verdict, stored assessment report, and redeemable report token that networks, issuers, and processors can fetch from the Trust Authority for disputes, liability, and oversight.

Improves fraud controls, decisioning, liability, and dispute handling.

consult hyperion

What is ATLAS the new open standard ?

The ATLAS Protocol Specification defines an open, interoperable framework for independently assessing AI-agent behavior in commerce. It describes the roles, message flows, evidence, authorization, privacy controls, and signed trust signals needed to evaluate whether an agent followed the user's intent and applicable policies. ATLAS is designed to complement agent identity and payment-authorization standards such as KYA-OS, AP2, and Verifiable Intent, while keeping sensitive user data off the payment rail.

https://hyperlab-fime.github.io/ATLAS-PUBLIC/atlas-protocol-specification-0.1draft.html

The ATLAS FAQ explains how an independent Assessor can evaluate an agent's behavior, verify whether it followed the user's intent and relevant rules, and produce a compact, signed trust signal for merchants, wallets, and payment networks.
It also shows how ATLAS complements emerging standards such as KYA-OS, AP2, and Verifiable Intent—without exposing the user's private conversation.

https://github.com/HYPERLAB-FIME/ATLAS-PUBLIC/blob/main/FAQ.md

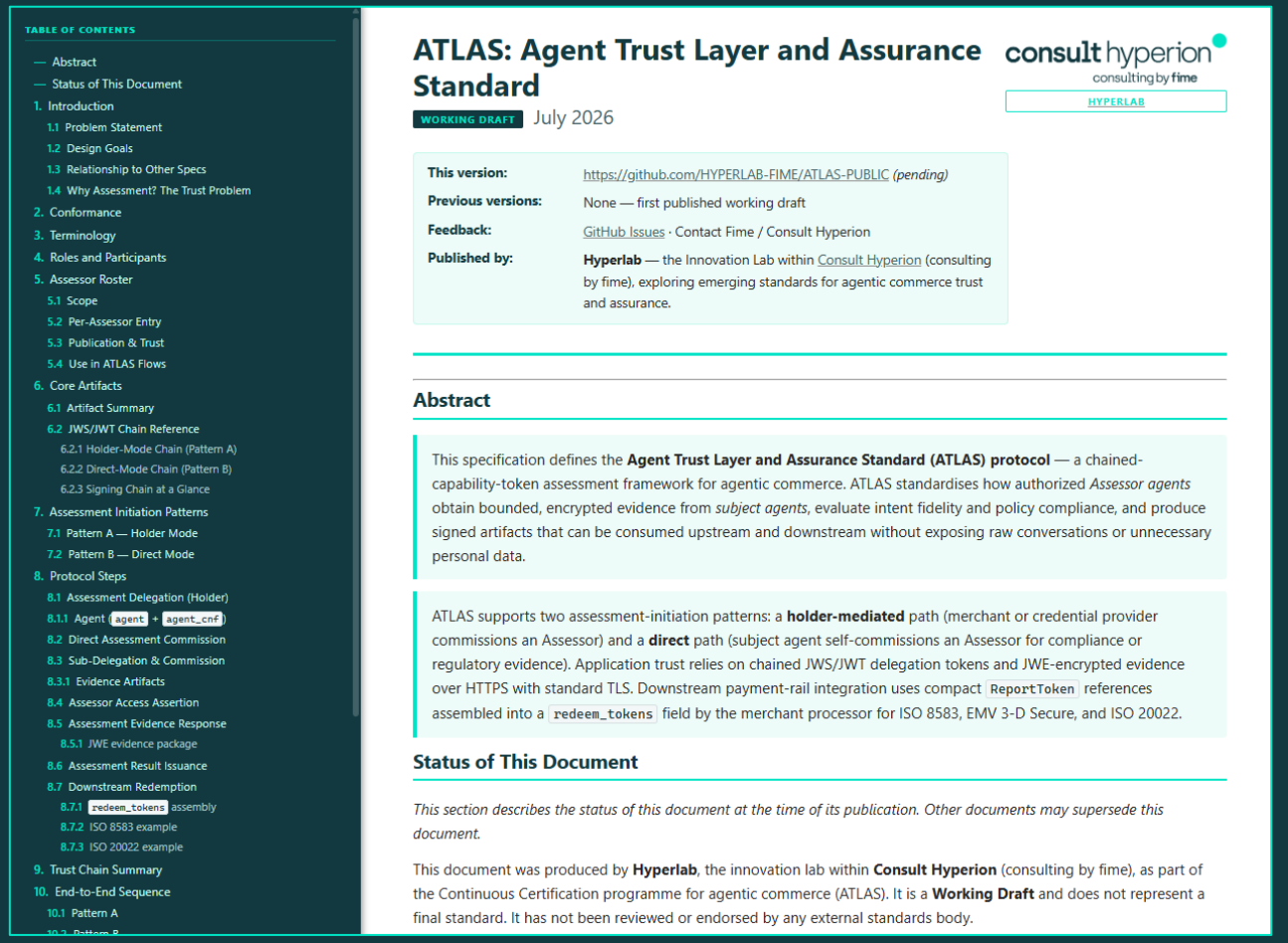


TABLE OF CONTENTS

- Abstract
- Status of This Document
- 1. Introduction
  - 1.1 Problem Statement
  - 1.2 Design Goals
  - 1.3 Relationship to Other Specs
  - 1.4 Why Assessment? The Trust Problem
- 2. Conformance
- 3. Terminology
- 4. Roles and Participants
- 5. Assessor Roster
  - 5.1 Scope
  - 5.2 Per-Assessor Entry
  - 5.3 Publication & Trust
  - 5.4 Use in ATLAS Flows
- 6. Core Artifacts
  - 6.1 Artifact Summary
  - 6.2 JWS/JWT Chain Reference
    - 6.2.1 Holder-Mode Chain (Pattern A)
    - 6.2.2 Direct-Mode Chain (Pattern B)
    - 6.2.3 Signing Chain at a Glance
- 7. Assessment Initiation Patterns
  - 7.1 Pattern A — Holder Mode
  - 7.2 Pattern B — Direct Mode
- 8. Protocol Steps
  - 8.1 Assessment Delegation (Holder)
  - 8.1.1 Agent (`agent` + `agent_cnf`)
  - 8.2 Direct Assessment Commission
  - 8.3 Sub-Delegation & Commission
  - 8.3.1 Evidence Artifacts
  - 8.4 Assessor Access Assertion
  - 8.5 Assessment Evidence Response
    - 8.5.1 JWE evidence package
  - 8.6 Assessment Result Issuance
  - 8.7 Downstream Redemption
    - 8.7.1 `redeem_tokens` assembly
    - 8.7.2 ISO 8583 example
    - 8.7.3 ISO 20022 example
- 9. Trust Chain Summary
- 10. End-to-End Sequence
  - 10.1 Pattern A

ATLAS: Agent Trust Layer and Assurance Standard

consult hyperion
consulting by fime
HYPERLAB

WORKING DRAFT July 2026

This version:	https://github.com/HYPERLAB-FIME/ATLAS-PUBLIC *(pending)*
Previous versions:	None — first published working draft
Feedback:	GitHub Issues · Contact Fime / Consult Hyperion
Published by:	**Hyperlab** — the Innovation Lab within Consult Hyperion (consulting by fime), exploring emerging standards for agentic commerce trust and assurance.

Abstract

This specification defines the **Agent Trust Layer and Assurance Standard (ATLAS) protocol** — a chained-capability-token assessment framework for agentic commerce. ATLAS standardises how authorized *Assessor agents* obtain bounded, encrypted evidence from *subject agents*, evaluate intent fidelity and policy compliance, and produce signed artifacts that can be consumed upstream and downstream without exposing raw conversations or unnecessary personal data.

ATLAS supports two assessment-initiation patterns: a **holder-mediated** path (merchant or credential provider commissions an Assessor) and a **direct** path (subject agent self-commissions an Assessor for compliance or regulatory evidence). Application trust relies on chained JWS/JWT delegation tokens and JWE-encrypted evidence over HTTPS with standard TLS. Downstream payment-rail integration uses compact `ReportToken` references assembled into a `redeem_tokens` field by the merchant processor for ISO 8583, EMV 3-D Secure, and ISO 20022.

Status of This Document

This section describes the status of this document at the time of its publication. Other documents may supersede this document.

This document was produced by **Hyperlab**, the innovation lab within **Consult Hyperion** (consulting by fime), as part of the Continuous Certification programme for agentic commerce (ATLAS). It is a **Working Draft** and does not represent a final standard. It has not been reviewed or endorsed by any external standards body.

consult hyperion

DEMO

consult hyperion

Thank you

consult hyperion

Discover more about how Consult Hyperion can help your business.

Email

jean-luc.dimanno@chyp.com

Making innovation possible.
Making the world work.

Consulting | Testing Solutions | Testing Services