The Cotrugli Ledger: A Truth and Evidence Layer for the AI Economy

NEO Accounting, NEO Cotruglian Triple Entry, and the Governance Architecture of Autonomous Agent Commerce

Dražen Kapusta[1], Matjaž Gams[2], Mario Brčić[3], Tali Režun[1]

[1] *COTRUGLI Business School, Zagreb, Croatia*
[2] *Jožef Stefan Institute, Department of Intelligent Systems, Ljubljana, Slovenia*
[3] *University of Zagreb, Faculty of Electrical Engineering and Computing, Zagreb, Croatia*

Corresponding author: Dražen Kapusta — drazen.kapusta@cotrugli.eu

April 2026

Abstract

Artificial intelligence needs at least a minimal truth layer, and one does not yet exist. In the next three to five years, most economic transactions on the planet will not be initiated by humans. They will be initiated by autonomous AI agents, robots, and software systems acting on behalf of humans, institutions, and other machines under conditions that are Networked, Exponential, and Orchestrated. The accounting and trust infrastructure of modern commerce was designed for human actors at human speed and will not survive contact with these conditions. Without a minimal truth layer, agents have no shared ground truth, every deployment implements its own logic, and the fragmentation that modern identity frameworks were designed to prevent is recreated at the transaction layer.

This paper presents the Cotrugli Ledger as the Truth and Evidence Layer that fills this gap. It is an accounting method designed specifically for autonomous agent commerce. Its operational primitive is the Policy-Aware Co-signed Receipt Object (PAC-RO), a cryptographically co-signed event object that binds mutual attestation between transacting parties at the moment of action, inseparable from its evidence and its governing policy. Its formal framework is NEO Cotruglian Triple Entry (NCTE). Its specified method for recognition eligibility under pinned policy is Policy-Bound Triple-Entry (PBTE). The method does not try to prove external reality. It records mutual attestation: what parties agreed at time T, under policy P, backed by evidence E. Once the attestation exists on the Truth and Evidence Layer, no party can silently alter it, no reconciliation is needed because there is no separate record to reconcile against, and downstream coordination proceeds on verifiable ground.

The Cotrugli Ledger is grounded in a philosophical tradition older than the discipline it extends. Ibn Khaldun in 1377 diagnosed the natural decay of civilizations from cohesion through luxury to senility, and named asabiyyah, social cohesion, as the mechanism that binds civilizations and whose failure unbinds them. Benedetto Cotrugli completed in Naples in 1458 a treatise arguing that honest records are the infrastructure that sustains long-distance cooperation where no court can enforce contracts and no army can collect debts. His work was printed in Venice in 1573 and published in complete English translation for the first time in 2016 by COTRUGLI Business School and Dražen Kapusta. Pacioli in 1494 codified the double-entry method that Cotrugli had described first in his earlier treatise, giving the technique its canonical mathematical form and making it teachable across the world. The tradition we extend is Cotruglian in its philosophical foundation and Paciolian in its mathematical codification, sitting within the Khaldunian diagnosis of institutional

decay that the accounting method is designed to interrupt.

We present the governance architecture of the Cotrugli Ledger in five components: a registry hierarchy for role and authority resolution; a twin scoring system (the Cotrugli Score for technical integrity, the Vanguard Score for behavioral reputation) that makes trust computable at machine speed and translates Khaldunian asabiyyah into a hard-coded metric; a three-stage dispute resolution model with deterministic arbitration; an AI autonomy framework (IAAF) with four authority levels, kill switches tied to score thresholds, and explainability artifacts; and a game-theoretic foundation that models honesty as the rational strategy in a large-network equilibrium. We engage adjacent 2025 literature on agent memory integrity, including MINJA, AgentPoison, and A-MemGuard, and position the accounting method as complementary to technical memory defenses rather than competitive with them. We place the discipline within the historical pattern of cross-border coordination standards such as ICAO, SWIFT, and JARUS that emerge through voluntary adoption. The research program is being validated at field scale through the Horizon Europe Vanguard AI consortium, sixteen partners across six EU Member States. The operational validation environment is SI-Chain, the first national blockchain infrastructure deployed at state scale anywhere in the world, powered by HashNET distributed ledger technology. The discipline is vendor-agnostic at the DLT substrate layer and model-neutral at the AI agent layer by design.

Keywords: *AI governance; triple-entry accounting; autonomous agents; policy-bound receipts; verifiable credentials; asabiyyah; constitutional AI; trust scoring; coordination standards; distributed ledger technology; Cotrugli.*

1. Introduction

1.1 The claim

Artificial intelligence needs at least a minimal truth layer, and one does not yet exist. This is the claim from which this paper proceeds. Autonomous AI agents are becoming economic actors at machine speed. Within three to five years, a material and strategically decisive share of economically consequential transactions will be initiated, mediated, or executed by AI agents rather than directly by humans. These transactions are carried out by robots, autonomous vehicles, industrial systems, and software orchestrators acting on behalf of humans, businesses, governments, and increasingly on behalf of other machines. The scale and timing of this transition were recognized in Kapusta and Liu [17], whose 2024 GENESIS proposal argued that the AI-to-AGI transition would require coordination on the ICAO model, operationalized through the United Nations Industrial Development Organization. Two years on, the specific substrate on which such coordination would rest has not been built. Agents have no shared ground truth to act upon. Every deployment implements its own accounting logic. The fragmentation that eIDAS 2.0 and comparable frameworks were designed to prevent at the identity layer is recreated at the transaction layer. Institutional decay, what Ibn Khaldun (1377) diagnosed as the failure of asabiyyah and what we call contextual drift in the digital economy, accelerates under automation rather than being mitigated by it.

The Cotrugli Ledger is the Truth and Evidence Layer that fills this gap. It is the accounting method designed for the minimum required substrate of an AI economy. Its operational primitive is the Policy-Aware Co-signed Receipt Object (PAC-RO), a cryptographically co-signed event object that binds mutual attestation between

transacting parties at the moment of action, inseparable from its evidence and its governing policy. Its formal framework is NEO Cotruglian Triple Entry (NCTE), named for the Adriatic merchant tradition from which double-entry bookkeeping emerged. The PBTE method (the formal specification of NCTE as it operates inside the Cotrugli Ledger) is developed in Kapusta and Brčić [18].

The method does not try to prove external reality. It records mutual attestation: what parties agreed at time T, under policy P, backed by evidence E. This is a weaker claim about the external world and a stronger claim about the coordination substrate, and it is the stronger claim that matters for the AI economy. Once the attestation exists on the Truth and Evidence Layer, no party can silently alter it, no reconciliation is needed because there is no separate record to reconcile against, and downstream coordination proceeds on verifiable ground. The Cotrugli Ledger produces clean records. Clean records are what autonomous agent commerce requires in order to function.

The Cotrugli Ledger sits within a philosophical tradition older than the accounting discipline it extends. Ibn Khaldun in 1377 diagnosed the natural decay of civilizations from cohesion through luxury to senility, naming asabiyyah, social cohesion, as the mechanism that binds civilizations and whose failure unbinds them [15]. Benedetto Cotrugli completed in Naples in 1458 a treatise arguing that honest records are the infrastructure that sustains long-distance cooperation where no court can enforce contracts and no army can collect debts. His work was printed in Venice in 1573 and published in complete English translation for the first time in 2016 by COTRUGLI Business School and Dražen Kapusta [7, 8]. Pacioli (1494) codified the double-entry method that Cotrugli had described first in his earlier treatise, giving the technique its canonical mathematical form and making it teachable across the world [25]. The tradition we are extending is Cotruglian in its philosophical foundation and Paciolian in its mathematical codification, sitting within the Khaldunian diagnosis of institutional decay that the accounting method is designed to interrupt.

Two modern extensions of the method have been proposed in the intervening centuries. Ijiri's momentum accounting [14a, 14b, 14c] introduced dynamic measurement of rates of change. Grigg's cryptographic triple-entry [13] reframed the third entry as a shared signed receipt binding counterparties. Both are genuine extensions. Both presuppose human actors. The emergence of autonomous AI agents as economic actors is a discontinuity of the same kind as the transitions that motivated each previous extension, and larger than either of them. This paper articulates the field-level claim that follows from the first author's three-year research program:

Claim 1 (NEO Accounting).

Autonomous AI agents as economic actors require a new accounting method in which the agent itself is the accounted entity. Every consequential action of an agent, including commercial commitments, settlements, acceptances, tool invocations, and writes to persistent memory, must be represented at event time by a co-signed, policy-bound, evidence-bearing event object. This object, not the subsequent ledger posting, is the primary unit of accounting truth. Ledgers, reports, analytics, and dispute resolution derive downstream of it. We call the resulting discipline NEO Accounting.

1.2 Scope of this paper

This is the third paper in a three-paper program. The civilizational argument, why accounting must be upgraded for the AI era, is developed in Kapusta [19]. The formal method specification, how PBTE operates as an Accounting State Machine with deterministic replay, fail-closed semantics, and three-lane governance, is developed in Kapusta and Brčić [18]. The present paper articulates the field-level claim: what discipline these two papers together constitute, how it positions against the adjacent 2025 literature, what governance architecture operationalizes it, and what remains to be validated.

We emphasize what this paper does and does not do. It does engage the convergent adjacent literature of 2025 on agent memory security, notably MINJA [9] on memory injection attacks and A-MemGuard [30] on retrieval-time defenses. It does not claim to subsume that work. It does not claim that NEO Accounting prevents memory corruption, adversarial attacks, or AI safety failures in any broad sense. It does claim, and defend, that the accounting method framing is distinctive, complementary to existing technical defenses, grounded in a six-hundred-year intellectual lineage, and the first articulation of a discipline with the features enumerated below. The accountant's claim is narrow and honest: we do not prevent the world from going wrong, we produce the records under which it remains inspectable, contestable, and repairable when it does.

1.3 The research arc

The first author has worked on this problem since 2023. The initial formulation grew from a specific intuition: in autonomous multi-agent systems operating on persistent memory, corrupted memories are indistinguishable from legitimate ones to the agent that holds them, and the reasoning consequences of a single corrupted memory can propagate across every subsequent retrieval that finds it analogically relevant. A defense that operates at retrieval time, while valuable, treats the symptom. The underlying problem is that memory writes, and more generally any consequential action by an autonomous agent, lack the properties that accounting discipline has historically provided for human commerce: bookability under explicit policy, structured evidence, co-signature, and first-class exception semantics.

From 2023 to 2025 the first author developed this framing in working papers and industry venues, including a triple-entry analysis of autonomous commerce that became the predecessor to PBTE. The framing was refined through discussion with the second author on global governance and coordination standard precedents [20], through discussion with the fourth author on explainable AI and trustworthy reasoning [10, 21, 22], and through discussion with the third author on longitudinal reasoning in ambient intelligence systems [11, 23]. The civilizational argument was published in Kapusta [19] in early 2026. The formal method specification was published in Kapusta and Brčić [18] shortly after, co-authored.

The operational validation environment of the research program has been SI-Chain, the first national blockchain infrastructure deployed at state scale anywhere in the world, powered by HashNET distributed ledger technology. SI-Chain and HashNET have served as the reference implementation substrate on which the Cotrugli Ledger has been built and tested across three years of research. The discipline itself is vendor-agnostic at the DLT substrate layer, a point we address precisely in Section 3. The choice of SI-Chain as reference substrate does not

constitute a claim that the Cotrugli Ledger requires any specific ledger; it documents the fact that the method has been exercised operationally, at national scale, on a DLT substrate meeting the requirements the method specifies.

During the same period, an independent literature emerged that validated parts of the problem space. MINJA [9], published March 2025 and subsequently accepted at NeurIPS 2025, demonstrated practical memory injection attacks against LLM agents with over 95% injection success rate through query-only interaction. AgentPoison [4] had demonstrated earlier that retrieval-augmented generation knowledge bases could be directly poisoned. MemoryGraft [26] extended this in late 2025 to persistent, session-independent behavioral drift. A-MemGuard [30] proposed a retrieval-time defense framework positioned as the first proactive defense framework for LLM agent memory. The broader survey by Torra [28] systematized the attack surface in early 2026. The fourth author of this paper continues active research into trustworthy memory architectures that protect memory systems from both attacks and mistakes, treating memory integrity as a first-order research problem in its own right.

This convergence is not coincidental. It reflects that the underlying problem, autonomous agents operating on persistent state whose integrity cannot be verified by the agent itself, is real, and that multiple research traditions have approached it from different angles. Our contribution is the accounting method framing, developed over three years and now formally specified. We regard the adjacent work as complementary evidence that the problem is genuine and as the backdrop against which our specific contribution must be precisely positioned.

1.4 Structure

The remainder of the paper proceeds as follows. Section 2 engages the adjacent bodies of work. Section 3 states the formal framework, including the four-layer architectural positioning. Section 4 addresses identity and standards alignment. Section 5 treats agent memory as the canonical application domain. Section 6 develops the philosophical foundation, tracing the Khaldunian, Cotruglian, and Paciolian lineage and establishing the game-theoretic basis of the NEO inversion. Section 7 presents the governance architecture in its five components. Section 8 sets out the research agenda and testable propositions. Section 9 develops the coordination layer pattern. Section 10 identifies the limits of the claim. Section 11 concludes.

2. Related Work and the Distinctive Move

Four distinct bodies of work are adjacent to the claim made in this paper. We address each in turn.

2.1 Triple-entry accounting as extended measurement

Ijiri's triple-entry bookkeeping [14a, 14b, 14c] was motivated by the observation that classical double-entry captures stocks and their changes but does not natively capture the force producing change, what Ijiri called momentum. The third entry in Ijiri's framework records rates of change, enabling dynamic measurement. This is a genuine extension of accounting method and established the important principle that accounting primitives are not fixed: new economies may require new primitives.

NEO Accounting inherits this principle. It differs in what the third entry denotes. In Ijiri, the third entry is a measurement of momentum within a single enterprise's accounts. In NEO Accounting, the third entry is the shared, co-signed, policy-bound event object that exists outside either party's unilateral representation, closer to Grigg's framing below, but extended to include explicit signature policy, structured evidence, and first-class exception semantics, and extended again to cover autonomous agents as signatories operating under bounded mandate credentials.

2.2 Triple-entry as cryptographic receipt

Grigg [13] proposed that a cryptographically signed receipt shared by counterparties should be treated as the authoritative record of a transaction, strengthening non-repudiation and evidentiary quality for inter-party commerce. This line of work influenced subsequent development of cryptographic audit trails and indirectly shaped blockchain-based accounting proposals [2, 6].

Grigg's proposal was designed for human commerce. Its implicit actor model assumes that signatories are humans or organizations represented by humans, that evidence is ancillary rather than structurally integral, and that disputes are resolved through mechanisms external to the receipt itself. NEO Accounting differs in three specific ways. First, it assumes autonomous agents as first-class signatories, with their authority bounded by mandate credentials verified at signing time. Second, it makes evidence structurally integral: the PAC-RO is incomplete without its evidence bundle. Third, it treats exceptions (disputes, overrides, partial acceptance, corrections) as first-class object types linked to the original event rather than as out-of-band communications. Kapusta and Brčić [18] formalize these differences in the PBTE Accounting State Machine.

2.3 Cryptographic provenance and agent commerce protocols

A recent and rapidly growing literature addresses cryptographic provenance for AI agent actions. Work on trusted execution environments and remote attestation for AI inference produces verifiable receipts that bind model outputs to the specific code and environment that generated them. The Sui Verifiable AI Control Plane [27] combines Walrus, Seal, and Nautilus for data, policy, and attestation. MAPLE [24] provides an agent operating system with WorldLine receipts. Agent commerce protocols, including Google's AP2 [12], Coinbase and Cloudflare's x402 [3, 5], and the emerging ERC-8004 standard [10a], provide cryptographic primitives for agent-initiated transactions at protocol level.

A further recent effort is the Verifiable AI Provenance Framework (VAP), published as an IETF Internet-Draft in January 2026 [30a]. VAP defines requirements for producing evidentiary-grade decision trails for automated systems using existing IETF security technologies, without defining new cryptographic primitives. The authors of the draft explicitly scope VAP as addressing the problem of recording and verifying what decisions an automated system made, when, and based on what inputs, while placing broader governance concerns such as authorization policies, risk classification, and approval workflows outside the framework's scope. VAP and its domain profiles, including the VeritasChain Protocol for financial trading audit trails [30b], are peer standards-track work in the provenance layer of the stack, not competing accounting methods. The comparative positioning of this paper's framework against VAP and against classical accounting governance is developed in

Section 2.6.

This work is engineering-valuable and solves real problems. However, none of it is an accounting method in the strict sense. The receipts are cryptographic artifacts for verifiability, not accounting primitives with formal semantics. None of these systems specifies standardized classes of event objects with role-object compatibility matrices, signature policies, and exception semantics as first-class constructs. None connects to the intellectual lineage of bookkeeping. The governance properties that distinguish an accounting method (recognition eligibility, deterministic replay, fail-closed semantics on undeclared dependencies, lifecycle-governed state transitions, pre-registered falsification regimes) are not specified, because the contribution is not being framed as an accounting method.

The distinction matters because accounting methods, historically, have been characterized by being repeatable, governed, and teachable. A cryptographic audit trail is an input to an accounting method; it is not one. NEO Accounting is an accounting method in this sense, and PBTE is its formal specification. Any cryptographic provenance system can be an implementation substrate for NEO Accounting, and we regard Sui, MAPLE, AP2, x402, and ERC-8004 as potential substrates rather than competing methods.

2.4 AI memory security: MINJA, A-MemGuard, and the 2025 convergence

A third adjacent literature emerged forcefully during 2025, addressing memory poisoning in AI agents from the security angle. This literature is the most important to engage precisely, because it operates in the problem space closest to ours.

For the accounting discipline, an agent's memory writes are consequential actions because the agent's future commercial behavior, its acceptances, commitments, and settlements, depends on its accumulated state. A corrupted memory produces unreliable accounting entries downstream, even when each individual entry passes signature validation and evidence check, because the judgment embedded in the entry rests on false accumulated context. NEO Accounting does not claim to detect or prevent memory corruption. It provides the structural record under which the provenance of each memory, and therefore the evidentiary basis of each downstream entry, remains inspectable. Memory poisoning is therefore not only a security vulnerability; it is an accounting integrity problem, and the accounting response to it is not defense but clean records.

Dong et al. [9] (MINJA) demonstrated that practical memory injection attacks against LLM agents are feasible through query-only interaction, achieving over 95% injection success rate and over 70% attack success rate across diverse agents. MINJA was accepted to NeurIPS 2025 and is the foundational empirical demonstration that memory poisoning is a serious practical threat, not a theoretical one. Closely related work includes AgentPoison [4] on RAG knowledge base poisoning with direct system access, and MemoryGraft [26] on persistent session-independent behavioral drift through semantic imitation heuristics. The fourth author of this paper has noted that having good memory is important, but AI agents, like people, are very sensitive and cannot tell the difference between fake or implanted memories and real ones. Once input into the memory system, fake memories affect all the reasoning that triggers them, potentially causing a landslide. His ongoing research program therefore develops trustworthy memory architectures that protect the memory system from both attacks and mistakes. That research is

complementary to, but distinct from, the accounting method presented here.

Wei et al. [30] (A-MemGuard) is the most prominent defense framework in this space. The authors propose consensus-based validation and a dual-memory architecture that detect anomalies by comparing reasoning paths derived from multiple related memories. They report reductions in attack success rates of over 95% on benchmark attacks with minimal utility cost. They position their work as the first proactive defense framework for LLM agent memory.

We engage A-MemGuard carefully. A-MemGuard is a technical defense mechanism that operates at retrieval time. It detects poisoned memories by observing reasoning-path anomalies and adapts through lesson accumulation. It does not modify the agent's core architecture, does not require cryptographic infrastructure, and does not specify what properties a memory write should have had when it was originally stored. Its contribution is real and valuable within its specified scope. NEO Accounting operates at a different layer. It specifies what properties every write to an agent's persistent state must have. It is an accounting method, not a detection technique. Its contribution is to frame memory integrity, and more generally the integrity of every consequential agent action, as a bookable property under accounting discipline, rather than as a security property to be enforced by detection. Under NEO Accounting, an agent's memory write is a PAC-RO whose policy validity is verified at ingestion. A poisoned write that fails policy validation is rejected before it enters memory. A-MemGuard-style consensus validation can serve as one mechanism within the exception-handling layer of NEO Accounting. These are complementary contributions. Both are first of a different kind: A-MemGuard is the first proactive defense framework for LLM agent memory in the technical defense sense; NEO Accounting is the first articulation of accounting-for-AI as a governance discipline in the method level sense.

2.5 The distinctive move, stated precisely

To our knowledge, this is the first explicit articulation of accounting-for-AI as a governance discipline centered on policy-bound, evidence-bearing records for autonomous economic action. This work is among the first, and to our knowledge the first in this exact form, to connect AI auditability, policy-bound verifiable credentials, and accounting-native governance into a unified truth-and-evidence layer for autonomous commerce.

Claim 2 (Distinctive contribution).

NEO Accounting is the first articulation of accounting-for-AI as a governance discipline, in which (a) the autonomous agent is the accounted entity; (b) every consequential action of the agent is bookable under explicit pinned policy; (c) evidence is structurally integral to the event object, not ancillary; (d) exceptions are first-class typed object classes linked to their predecessors; (e) the ledger and all downstream reports derive from event objects rather than generating them; (f) recognition eligibility under pinned policy is a verifiable property of a shared receipt lifecycle computed by deterministic replay [18]; and (g) the method is grounded in the intellectual lineage of double-entry bookkeeping [7, 8, 25], momentum accounting [14b], and cryptographic triple-entry [13], extended to autonomous actors. No prior work combines these features. No prior work frames the contribution as an accounting method with formal lineage.

What is novel is the integration into a coherent accounting method with formal lineage, the framing of the agent as the accounted entity, and the articulation of this integration as a discipline.

2.6 Comparative positioning

The claim of Section 2.5 is sharpened by comparing NCTE and PBTE explicitly against the two frameworks whose scope overlaps most directly with autonomous agent commerce: the classical accounting, audit, and governance architecture consolidated in COSO and IAASB materials, and the emerging IETF VAP framework for AI provenance. Table 1 presents the comparison across ten dimensions.

Table 1. Comparative positioning of classical accounting, audit, and governance; IETF VAP; and NCTE / PBTE as framed in this paper.

Dimension	Classical accounting / audit / governance	IETF VAP	NCTE / PBTE (this paper's framing)
Primary problem addressed	Reliable financial reporting, internal control, stewardship, and assurance over organizational activity	Evidentiary-grade provenance and verifiable decision trails for AI or automated systems	Making autonomous AI action economically recordable, policy-bound, auditable, and governable
Core object	Ledger entry, voucher, transaction record, financial statement, audit evidence	Provenance record, decision trail, verifiable event trail	Policy-bound, evidence-bearing AI transaction object or receipt (PAC-RO)
Primary assumed actor	Human-managed organization, management, accountant, auditor	AI system, automated decision pipeline, or algorithmic actor	Autonomous AI agent acting under delegated authority in commerce
When evidence is created	Often recorded operationally, then reconciled, reviewed, and audited later	Captured as verifiable provenance around decision events	Bound to the economic action itself at the time of execution
Audit logic	Internal controls plus external assurance over reporting and evidence sufficiency	Independent verification of integrity, provenance, completeness, and accountability of AI trails	Accounting-native audit trail for autonomous action, including policy basis, evidence links, and dispute handling
Governance locus	Board, management, audit committee, internal control, external auditor	Cross-organizational accountability for AI decisions and evidence trails	Transaction-level governance of machine-initiated commerce
Main technical substrate	ERP systems, ledgers, control frameworks, audit files	Existing IETF security technologies and adjacent transparency services	Verifiable Credentials plus truth-and-evidence layer plus policy-bound receipt object plus multi-tier governance logic
What it does best	Financial accountability and assurance over organizational reporting	Cryptographically verifiable provenance for AI decisions and workflows	Connects autonomous action to accounting treatment, auditability, and governance in one object

Dimension	Classical accounting / audit / governance	IETF VAP	NCTE / PBTE (this paper's framing)
What it does not natively solve	Native treatment of machine-initiated economic activity at agent scale	Economic booking logic, accounting classification, or commercial settlement semantics	Requires standards mapping, institutional adoption, and rigorous operational profiles
Best novelty claim	Established baseline	Strong adjacent provenance framework	Accounting-native governance method for autonomous AI commerce, not provenance alone

Note. Table note. The classical column is anchored in the mainstream architecture of internal control, audit evidence, governance, and assurance reflected in COSO and IAASB materials. The VAP column is anchored in the January 2026 IETF Internet-Draft draft-kamimura-vap-framework-00, which describes VAP as an architectural framework for evidentiary-grade AI decision trails using existing IETF security technologies, together with the IETF SCITT (Supply Chain Integrity, Transparency, and Trust) architecture on which VAP builds and with which it interoperates. The NCTE / PBTE column presents this paper's own proposed framing, not an already adopted standard.

The comparison makes two things explicit. First, classical accounting governance and IETF VAP are not competitors to NCTE; each addresses a problem space that NCTE does not, and each leaves unaddressed the problem space that NCTE fills. Classical frameworks are the baseline architecture of organizational financial accountability; they do not natively treat machine-initiated economic activity at agent scale. VAP is a strong adjacent provenance framework for AI decision trails; by the authors' own scoping, it does not address economic booking logic or accounting classification. NCTE and PBTE together propose the accounting-native governance method that neither of the other frameworks provides, while depending on both for substrate: VAP and comparable transparency architectures for cryptographically verifiable provenance, and classical accounting systems for downstream statutory reporting.

3. Formal Framework

We summarize the formal framework at the level appropriate for field-level positioning. The full specification appears in Kapusta and Brčić [18]. Before proceeding, we note the relationship between the two names used across the paper. The Cotrugli Ledger is the name we use for the deployed artifact and for the conceptual identity of the discipline as a whole, including in this paper's title and throughout general discussion. NEO Cotruglian Triple Entry (NCTE) is the name of the technical framework that specifies the method inside the Cotrugli Ledger. Where we refer specifically to the method-level specification, we use NCTE; where we refer to the deployed artifact or its conceptual identity, we use the Cotrugli Ledger. Section 3 and Section 7 use NCTE most frequently because they treat the method-level specification directly. The broader name, NEO Accounting, refers to the scholarly discipline at field level of which the Cotrugli Ledger and NCTE are specific instances.

3.1 The Truth and Evidence Layer and its architectural position

Definition 1 (Truth and Evidence Layer).

The Truth and Evidence Layer, which we also call the Cotrugli Ledger in its operational form, is the data and governance layer that sits above the integrity substrate provided by distributed ledger technology and below the enterprise accounting systems that consume its outputs. It accepts, validates, and persists event objects at event time, and produces queryable, auditable, recognition-eligible representations of agent actions. The Layer requires a DLT substrate as its integrity anchor but is vendor-agnostic at that layer: any distributed ledger providing tamper-evidence, cryptographic co-signature support, credential binding, and deterministic timestamping may serve. The research program reported in this paper has used SI-Chain, powered by HashNET distributed ledger technology, as the reference implementation substrate.

The Layer occupies a specific position in a four-layer architectural stack. At the bottom sits the identity and credential layer, which provides role credentials, mandate credentials for autonomous agents, and the cryptographic primitives required to bind signatures to identifiable principals. The W3C Verifiable Credentials Data Model v2.0 [29] and the EU Digital Identity Wallet framework [14d] are the canonical instances of this layer. Above the identity layer sits the DLT integrity substrate. Any ledger providing the required guarantees may serve in this role. In the research program reported here, the substrate is SI-Chain, a national blockchain infrastructure deployed at state scale and powered by HashNET distributed ledger technology. Above the DLT substrate sits the Truth and Evidence Layer itself, which is the layer specified in Definition 1 and the subject of this paper. It provides recognition eligibility gating, policy-bound accounting treatment, structured evidence validation, and lifecycle state management for the event objects it persists. Above the Truth and Evidence Layer sit the enterprise accounting systems, which include ERP platforms, statutory reporting pipelines, tax systems, and general ledgers. These systems consume the PAC-ROs produced by the Truth and Evidence Layer and derive their downstream journal entries, reports, and filings from them.

Data flows upward through the stack. Identity credentials flow upward into the DLT substrate at signing time. DLT-anchored co-signed receipts flow upward into the Truth and Evidence Layer where they undergo policy evaluation and recognition gating. Recognition-eligible PAC-ROs flow upward into the enterprise accounting systems where they map, through deterministic transformation, into one or more journal entries. Each layer uses the one below it as infrastructure and feeds the one above it with its outputs.

This architectural positioning is the critical distinction. NCTE requires a DLT substrate beneath it to deliver its structural guarantees of tamper-evidence, cryptographic co-signature, and credential-bound authority. Without the substrate, the PAC-RO collapses into a signed document under implicit governance and loses the properties that distinguish it from conventional accounting records. NCTE leaves existing enterprise accounting systems above it untouched. Enterprises can adopt NEO Accounting without replacing their ERP systems, statutory reporting pipelines, or tax infrastructure, because the method produces outputs that map directly into the journal entry conventions those systems already use. The Layer completes the stack. It does not compete with any part of it.

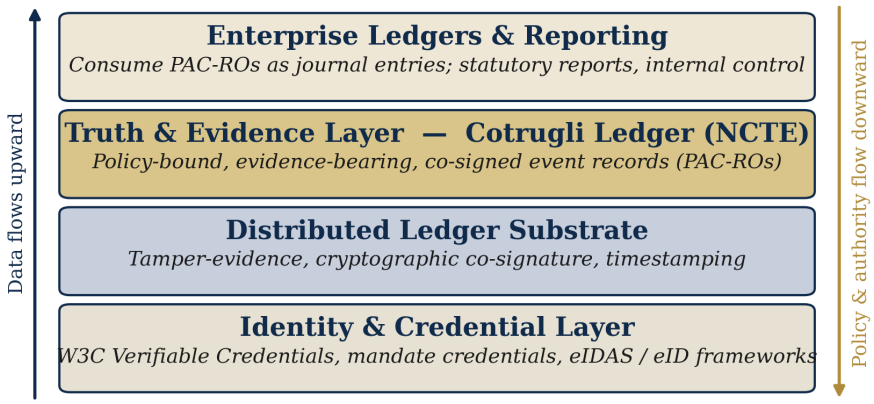

Figure 1. The four-layer architecture of agent commerce. NCTE occupies the Truth and Evidence Layer.

Figure 1. The four-layer architecture of agent commerce. NCTE occupies the Truth and Evidence Layer, with a DLT integrity substrate beneath it and enterprise reporting systems above.

3.2 The Policy-Aware Co-signed Receipt Object

Definition 2 (PAC-RO).

A Policy-Aware Co-signed Receipt Object consists of four inseparable elements: Facts, the minimal canonical data describing the event; Evidence, a structured bundle of attachments, attestations, source references, and sensor telemetry that substantiates the facts; Signature Policy, an explicit specification of who must sign by role, credential, or mandate, at what thresholds, and under what preconditions; and Exception Semantics, a first-class typed representation of disputes, overrides, partial acceptance, variances, and corrective events, structurally linked to the original PAC-RO. A PAC-RO is valid only if all four elements are present and coherent.

The PAC-RO does not try to prove external reality. It proves mutual attestation. Rather than certifying that a shipped box definitively contains gold, the PAC-RO certifies that all relevant parties agreed and attested, at a specific moment, to a shared set of data points about the box, under a specific signature policy, backed by a specific evidence bundle. This mutual attestation becomes the unassailable foundation for subsequent coordination, because no party can unilaterally alter what all parties agreed, and no reconciliation is required because there is no separate record to reconcile against.

The inseparability of the four elements is essential. A cryptographic receipt without explicit policy is not a PAC-RO; it is a signed artifact whose governance is implicit and therefore unverifiable. A policy specification without structured evidence is not a PAC-RO; it asserts conditions without ability to verify them. Exceptions without first-class semantics revert to narrative reconstruction, which is the failure mode the method is designed to eliminate. The PBTE specification [18] extends the PAC-RO to a seven-field minimum binding with an Accounting State Machine gating recognition eligibility through five lifecycle states.

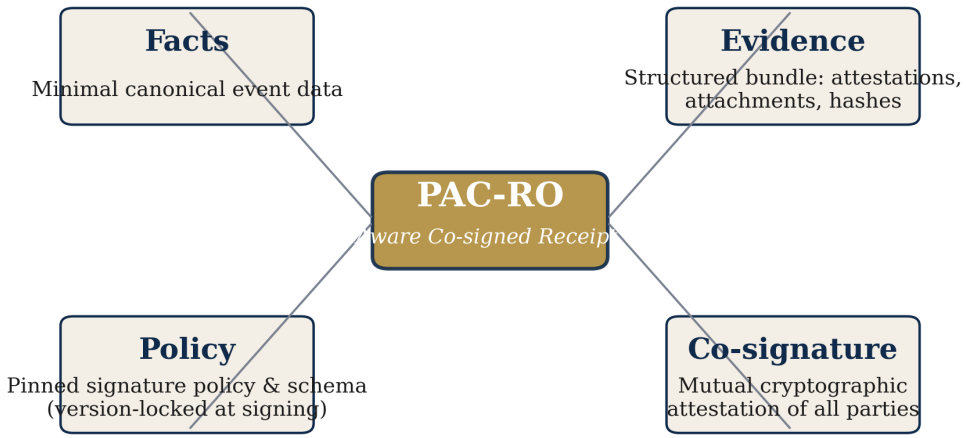

Figure 2. The PAC-RO. The four elements are inseparable; removing any one breaks the receipt.

Figure 2. The PAC-RO. Facts, Evidence, Policy, and Co-signature are inseparable; removing any element breaks the receipt.

3.3 Standardized event classes

NEO Accounting specifies five invariant event object classes: Commitment, Fulfillment, Acceptance, Settlement, and Exception. Industries may specialize schemas and policies within each class; the class taxonomy itself is invariant across domains. This invariance is what allows the method to be recognizable as the same discipline across industries, analogously to how double-entry journal entries for rent and for wages are instances of a common abstraction. Shared semantics across the network are a load-bearing property of the method: different agents, regardless of their internal technology stacks or training, share an identical understanding of what a transaction or agreement entails because they use the same PAC-RO schema.

3.4 Relationship to downstream ledgers

NEO Accounting does not replace enterprise ledgers. Every PAC-RO maps, through deterministic transformation, to one or more journal entries in one or more downstream ledgers. Multiple ledgers may consume the same PAC-RO. The ledgers remain the authoritative record for statutory reporting purposes. The PAC-RO is the authoritative record for what happened, why it was authorized, and under what evidence. Enterprises can therefore adopt NEO Accounting without replacing existing ERP, accounting, or statutory reporting pipelines.

4. Identity, Authority, and Standards Alignment

NEO Accounting is not an identity system. It consumes identity and authority claims from external standards and uses them as inputs to signature policy evaluation. The W3C Verifiable Credentials Data Model v2.0 [29] supplies the canonical format for role and mandate credentials. The EU Digital Identity Wallet ecosystem [14d] supplies the regulatory context within which credentials are issued, held, and

revoked.

For autonomous agents, the critical primitive is the mandate credential: a bounded authority credential specifying scope, limits, and decision-receipt requirements. An agent operating under a mandate credential may sign PAC-ROs within its scope; exceeding the mandate triggers escalation to a human signatory. This integration is the bridge between accountable autonomy and the current European regulatory trajectory, and it is extended within the Cotrugli Ledger's technical framework through the Boundary Credential Verifiable Credential (PAC-BCVC), the explicit, cryptographically bound delimitation of what an agent is authorized to do under a given mandate. We return to this in Section 7 where we present the AI autonomy framework.

5. Agent Memory as the Canonical Application

Although NEO Accounting applies uniformly to consequential agent actions, agent memory is the canonical application domain and the one most directly connected to the 2025 security literature. An LLM agent with persistent long-term memory has a specific epistemic vulnerability. Unlike a single-shot inference where the context is ephemeral and inspectable, a memory-augmented agent retrieves from an accumulated store built over weeks, months, or years. A false episode entering that store, through adversarial injection, sensor error, miscalibrated inference, direct poisoning, or semantic grafting, does not cause only one bad answer. It contaminates every future retrieval that finds it analogically relevant. Contrastive retrieval, which most production memory systems use, makes this worse because it is explicitly designed to surface analogous patterns.

This is the memory landslide problem. The agent has no native meta-cognitive alarm saying how do I actually know this? Memory to the agent is just retrieval results. There is no native notion of this memory has weaker evidence than that one or this memory's provenance is contested. The problem is sharpened in longitudinal ambient intelligence settings, where memory accumulates over months or years and the integrity of each write compounds into the integrity of the whole store [11, 23].

Under NEO Accounting, every memory write is a PAC-RO. The facts describe the episodic content, the evidence bundle links the write to its originating observation or inference, the signature policy specifies the mandate credential under which the agent is authorized to write, and the exception semantics provide for correction, retraction, and dispute of prior writes. If all writes are PAC-ROs under a given signature policy, three structural affordances follow. Writes that fail policy validation at ingestion cannot enter the memory store. Retrieval operations surface provenance alongside content, making the evidentiary basis of every retrieved memory inspectable. Retroactive re-validation of past writes against updated trust state becomes possible through evidence replay.

The accounting discipline does not prevent memory corruption and does not detect novel attacks. It produces the structural record under which corruption, when it occurs, is inspectable and contestable. Retrieval-time defenses such as A-MemGuard catch what slips through at retrieval; accounting discipline at the write boundary makes ingestion-time rejection and retroactive detection tractable. Neither replaces the other, and a serious deployment would use both, with the specific collaboration between method-level accounting and technique-level defense treated

as an empirical question to be resolved with the security research community.

6. Philosophical Foundation: Asabiyyah, Cotrugli, and the NEO Inversion

6.1 Ibn Khaldun and the civilizational diagnosis

In 1377, Ibn Khaldun completed the Muqaddimah, the introduction to his universal history [15]. At its center is a diagnostic claim about how civilizations cohere and how they decay. The mechanism of cohesion he named asabiyyah: the social bond, the group feeling, the shared sense of obligation that allows a community to act in concert and to extend trust beyond the range of immediate personal relationship. Asabiyyah is what allows cooperation to scale. Its decay is what causes civilizations to fragment.

Khaldun's diagnosis was that civilizations follow a natural decay cycle. An early generation, tested by necessity, builds strong asabiyyah and the institutions its cohesion sustains. A middle generation inherits the institutions and begins to enjoy the luxury they produce. A later generation loses the cohesion, loses the discipline, and loses the ability to defend what was built. Institutions that once functioned through shared norms become dependent on coercion, then cease to function at all. This is institutional senility. Khaldun saw it as inevitable for any civilization that cannot refresh its asabiyyah.

The digital economy exhibits a precise analog of this decay, which we call contextual drift. Humans, organizations, and AI models maintain their own siloed ledgers of truth, producing different and unverifiable versions of reality. Reconciling these separate records is slow, error-prone, and creates friction that kills coordination at scale. AI deployed into these fragmented structures automates and accelerates the mistrust rather than mitigating it. The failure mode Khaldun diagnosed at civilizational time scales now occurs at machine speed. The question this paper poses is whether the accounting discipline, correctly extended, can interrupt the decay rather than accelerate it.

6.2 The Cotruglian response

Eighty-one years after Khaldun, Benedetto Cotrugli completed in Naples a treatise that proposed part of the answer [7, 8]. The merchant, Cotrugli argued, operates across borders, languages, religions, and jurisdictions where no court can enforce his contracts and no army can collect his debts. What travels with the merchant is his name. If his name is honest, his bills of exchange are accepted in cities he has never visited by men who have never met him. If his name is dishonest, even his own correspondents will not extend him credit. Honesty is not a moral luxury. It is the only infrastructure that scales across boundaries which armies and laws cannot.

The accounting chapter in Della mercatura sits inside a four-book argument about what kind of person the merchant must be for commerce to work at all. Pacioli (1494) codified the double-entry method that Cotrugli had described first in his earlier treatise, giving the technique its canonical mathematical form and making it teachable across the world [25]. The tradition the accounting method carried forward was Cotruglian in foundation and Paciolian in form. Both contributions are essential. The Cotruglian philosophical foundation explains why the technique works, what kind of social infrastructure it supports, and why it has persisted for

more than five centuries. The Paciolian mathematical codification explains how the technique travels, how it is taught, and how it becomes a discipline rather than a local practice.

Fifty-five years after Cotrugli completed his manuscript, Machiavelli published Il Principe and reached the opposite conclusion. The world is hostile, men are wicked, the survivor simulates virtue while practicing whatever cruelty the moment requires [20a]. Trust is for fools. For most of the subsequent five centuries, Machiavelli has been operationally winning the argument even when Cotrugli has been ethically winning it. The reason is structural. In a world where commerce moves at the speed of sailing ships, the defector who extracts value and moves on faster than the news travels profits before reputation catches up. The information velocity favored defection. Honest merchants were right about the long run, but defectors took the short run, and enough short runs compounded into careers.

6.3 The NEO inversion and its game-theoretic foundation

Three features of the NEO economy collapse this asymmetry for the first time. Defection becomes detectable in seconds rather than seasons, because every PAC-RO is a permanent cryptographic witness. The network effect of trust compounds at machine speed, because an agent accumulating honest receipts across millions of micro-interactions accrues a reputation that functions as an economic license. Principals cannot be operationally separated from the agents they deploy, because mandate credentials are cryptographically linked to the principal. Cynicism becomes its own confession. The arithmetic that favored defection at human scale inverts at machine scale.

The inversion can be stated game-theoretically. In a consortium network of meaningful size, honesty becomes the rational strategy when the expected cost of detection exceeds the expected gain from defection. Formally, the equilibrium holds when P(detection) multiplied by Loss(reputation plus exclusion plus penalties) exceeds Gain(collusion). In a large network, P(detection) approaches 1 because the observation surface grows with the number of participants and the audit trail is immutable. Loss grows with the value of continued membership, which at consortium scale becomes substantial because exclusion from a shared infrastructure imposes costs far beyond any single transaction. Gain from collusion is bounded by the size of the specific transaction being manipulated. We have modeled this equilibrium on a projected network of 700 participating firms, a scale chosen to illustrate consortium-grade deployment dynamics rather than to claim empirical validation. At that scale, in the projected model, honesty becomes the rational strategy by a wide margin. The Cotrugli Ledger makes P(detection) infrastructurally high, and the Vanguard Score (Section 7) makes Loss(reputation plus exclusion) infrastructurally visible. The mathematics of machine-speed commerce favors the merchant over the prince, for the first time in five centuries, not because humans have become better but because the architecture has changed.

We call the resulting discipline NEO Cotruglian because Cotrugli named its philosophical foundation first and because the merchant tradition he described is the only ethical operating system we have seen that scales to a network of billions of heterogeneous agents. The discipline interrupts the Khaldunian decay cycle at its root: it transforms asabiyyah from a soft variable subject to generational erosion into a hard-coded, cryptographically anchored metric that cannot be lost through

complacency. This is not a nostalgic return to pre-modern ethics. It is the recognition that the infrastructure we are about to build is finally sufficient to make Cotrugli's thesis operationally binding rather than aspirational, and to make Khaldun's diagnosis a design input rather than a historical inevitability.

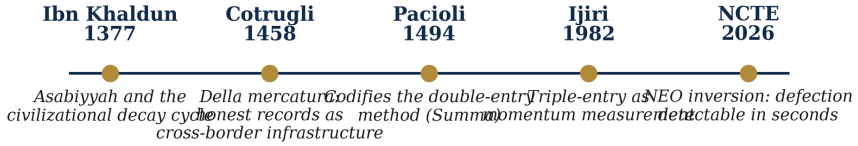

Figure 6. Lineage of the method. NCTE extends a tradition that begins with Khaldun's diagnosis and Cotrugli's response.

Figure 3. The lineage of NEO Cotruglian Triple Entry. The method extends a tradition reaching back to Khaldun and Cotrugli.

7. The Governance Architecture of the Cotrugli Ledger

The formal framework of Section 3 and the philosophical foundation of Section 6 together specify what NEO Accounting is and why it works. Making the discipline operational at consortium scale requires a governance architecture, the set of institutional mechanisms by which authority is resolved, reputation is measured, disputes are adjudicated, and autonomous agents are bounded. We present the architecture in five components, each developed at the depth appropriate to a field-level paper. The full formal specification of each component is companion work.

7.1 Registry hierarchy

Authority and role resolution under NCTE are handled through four primary registries arranged in a strict precedence hierarchy. The Governance Council Registry (GCR) is authoritative for governance voters and their voting weights. The Arbiter Registry (AR) manages human arbiters and their domain and technical qualifications. The Trusted Issuers Registry (TIR) is authoritative for regulated entities such as Qualified Trust Service Providers under eIDAS 2.0, auditors, and carbon attesters. The Party and Role Registry (PRR) handles enterprise decentralized identifiers, roles, and delegation relationships.

When a signer's role or authority is disputed, the verifier follows a strict hierarchy: GCR takes precedence over AR, AR over TIR, TIR over PRR. This ordering reflects the institutional weight of each registry: governance council authority trumps arbiter authority, which trumps trust service provider status, which trumps enterprise-level role assertion. The hierarchy produces deterministic dispute resolution for role questions without requiring narrative adjudication.

The registry architecture is compatible with the W3C Verifiable Credentials Data Model and integrates with national and regional identity ecosystems through TIR, which inherits trust from the Qualified Trust Service Provider framework in jurisdictions where that framework exists. In jurisdictions without a mature trust service framework, TIR is populated by alternative credential issuance mechanisms meeting equivalent standards.

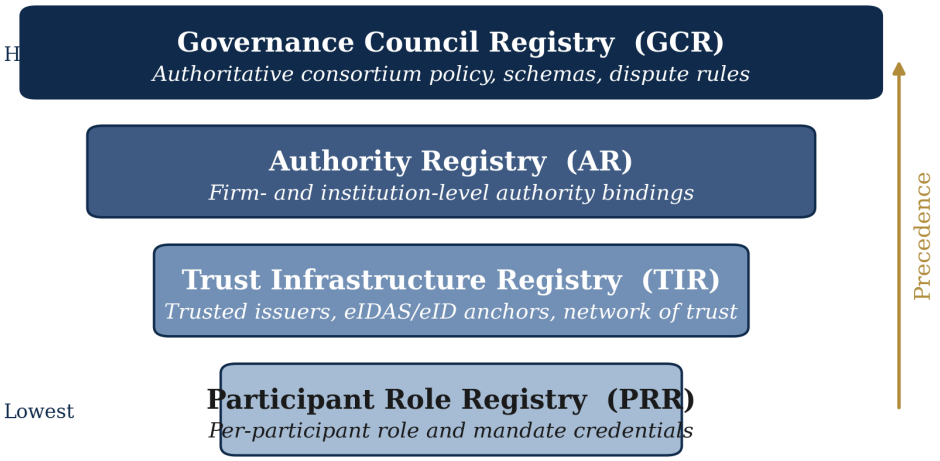

Figure 3. Registry hierarchy under NCTE. Conflicts resolve top-down: GCR > AR > TIR > PRR.

Figure 4. Registry hierarchy under NCTE. Disputes over a signer's authority resolve top-down: GCR > AR > TIR > PRR.

7.2 The twin scoring system

Trust under NCTE is not a soft social variable. It is measured empirically through two distinct scoring layers, each bounded between 0 and 100, and each computed from cryptographically verifiable on-ledger evidence.

The Cotrugli Score measures technical integrity. It tracks metrics such as proof of agreement consistency (the percentage of co-signed receipts that validate cleanly), schema compliance (the extent to which an agent's PAC-ROs conform to the standardized event classes), and anchoring reliability (the timeliness and completeness of DLT anchoring). The Cotrugli Score is the accounting-technical measure of whether an agent or firm operates within the discipline. It is closer to a credit score for transactional integrity than to a reputation score in the conventional sense.

The Vanguard Score measures behavioral reputation. It tracks dispute performance (win and loss rates in the three-stage dispute resolution model), corrective action compliance (the completeness and timeliness of remedies following a resolution), and cooperation reliability (behavior across repeat interactions with diverse counterparties). The Vanguard Score is the direct digital translation of Ibn Khaldun's asabiyyah. Khaldun treated social cohesion as a soft variable that civilizations could not measure, could not transmit across generations, and therefore could not defend against decay. The Vanguard Score makes asabiyyah measurable, transferable, and defensible by anchoring it in verifiable behavioral evidence on the ledger. What Khaldun diagnosed as a civilizational vulnerability becomes, under NCTE, a continuously observable metric that the governance system can act upon.

The two scores work in tandem. The Cotrugli Score measures whether an agent follows the rules of the discipline. The Vanguard Score measures how the agent behaves under stress, particularly when disputes arise or commitments fail. A high Cotrugli Score with a low Vanguard Score indicates an agent that technically complies but cooperates poorly. A low Cotrugli Score with a high Vanguard Score

indicates an agent that acts honorably but does not yet meet the technical standards of the discipline. Both scores are needed for reliable assessment, and the dispute resolution model of Section 7.3 and the autonomy framework of Section 7.4 use them together.

7.3 Three-stage dispute resolution

Disputes under NCTE are handled through a deterministic three-stage arbitration path that does not rewrite history. Each resolution is appended to the ledger as a Resolution Attestation (PAC-RA) linked to the original PAC-RO, preserving the immutable predecessor and making the governance trajectory fully auditable.

Stage one is automated triage, resolving within zero to twenty-four hours of the disputed event. An adjudicator bot performs evidence consistency checks, verifies signature policy compliance, and flags clear cases for accelerated resolution or escalation. Most micro-disputes resolve at stage one because the evidence bundle and signature policy embedded in the PAC-RO provide sufficient basis for automated judgment.

Stage two is expert arbitration, resolving within twenty-four hours to fourteen days. A rotating panel of three members (one technical expert, one domain expert, and one neutral Qualified Trust Service Provider or equivalent) reviews escalated cases and issues a binding resolution. Panel composition is drawn from the Arbiter Registry, and the rotation prevents capture of the arbitration process by any single institutional interest.

Stage three is consortium governance, resolving within up to thirty days. Systemic issues, appeals from stage two, or disputes affecting consortium-wide rules are adjudicated by an M-of-N vote of consortium members drawn from the Governance Council Registry. For systemic risks, a two-thirds majority is typically required. The consortium governance stage is the mechanism by which the rules of the discipline itself can evolve, and by which the consortium responds to novel situations that the existing schema and policy do not anticipate.

The dispute resolution mechanism is itself a governance asset. The Vanguard Score ingests dispute outcomes as behavioral evidence. An agent or firm that repeatedly escalates disputes that it loses accumulates negative Vanguard Score, which feeds back into its standing in future transactions and its authorized autonomy level under the IAAF framework of Section 7.4. The system is self-reinforcing: honest behavior produces receipts that compound into trust, and dishonest behavior produces disputes that compound into exclusion.

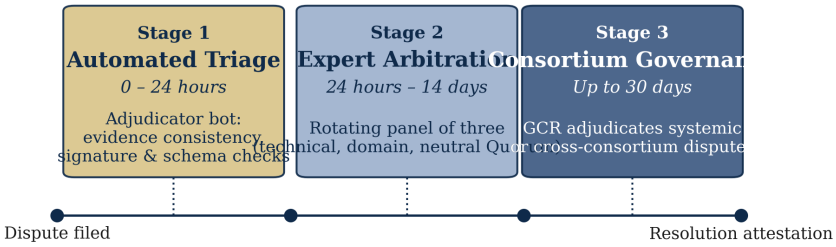

Figure 4. Three-stage dispute resolution. No stage rewrites history; resolutions are appended as Resolution Attestations.

Figure 5. Three-stage dispute resolution. No stage rewrites history; outcomes are appended as Resolution Attestations.

7.4 AI autonomy governance (IAAF)

The IPCE AI Agent Autonomy Framework (IAAF) manages the risks of autonomous systems by tying their authority to the parent firm's trust scores and to verifiable behavioral evidence on the ledger. IAAF specifies four autonomy levels for AI agents operating under NCTE.

Level 0 (L0) is Observer. The agent can read the ledger, analyze patterns, and produce recommendations, but cannot sign PAC-ROs. No commercial action is authorized. Level 1 (L1) is Co-signer. The agent can co-sign PAC-ROs within tightly bounded parameters, always alongside a human co-signer whose signature is required for validity. Level 2 (L2) is Delegated Agent. The agent can sign PAC-ROs autonomously within a pre-defined mandate scope, without human co-signature, but actions above a mandate threshold trigger automatic human notification and can be revoked within a defined window. Level 3 (L3) is Full Autonomous. The agent can sign PAC-ROs autonomously across the full mandate scope granted by its principal, subject to the signature policy and boundary credential attached to each action.

The movement between levels is not discretionary. An agent qualifies for higher autonomy levels only when the parent firm's Cotrugli and Vanguard scores exceed threshold values. If either score drops below its threshold, the system automatically downgrades the agent's autonomy level. This is a structural kill switch: a firm whose integrity or reputation degrades loses the ability to operate at high autonomy until it rebuilds its standing through verifiable behavior. The kill switch operates at machine speed, without requiring discretionary human intervention, and without requiring the parent firm's cooperation. This is what makes the system robust to the failure modes Khaldun identified at civilizational time scales.

Every agent action under IAAF must produce two explainability artifacts linked to the PAC-RO it generates. The Policy-Aware Co-signed Justification Chain (PAC-JC) is a machine-readable record of the decision factors, inputs, and reasoning steps that produced the action. The Policy-Aware Co-signed Boundary Credential Verifiable Credential (PAC-BCVC) is the cryptographically verifiable attestation that the agent acted within the scope of its boundary credentials at the time of action. Together, PAC-JC and PAC-BCVC make every autonomous action structurally inspectable and structurally bounded. An agent cannot validly sign a PAC-RO without producing both, and both are anchored to the ledger alongside the PAC-RO itself.

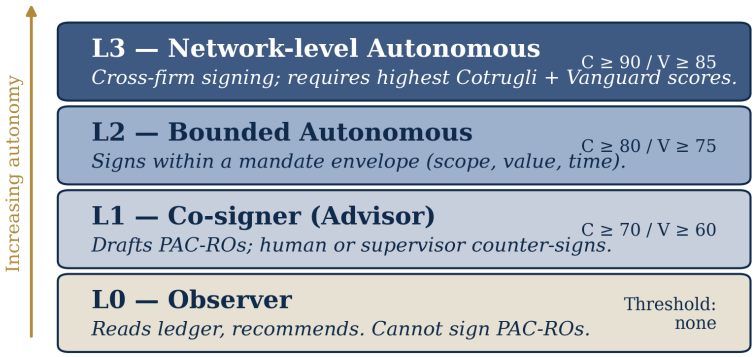

Figure 5. IAAF autonomy ladder. Movement between levels is gated by parent-firm Cotrugli and Vanguard scores.

Figure 6. IAAF autonomy ladder. Movement between levels is gated by the parent firm's Cotrugli and Vanguard scores.

7.5 Game-theoretic foundation

The four components above (registries, scoring, dispute resolution, autonomy framework) produce a system in which honest behavior is not only ethically preferable but rationally optimal. The governance architecture is underpinned by a game-theoretic equilibrium that becomes stable at consortium scale.

In the equilibrium, each participating firm faces a choice at every transaction: honest behavior, which produces clean PAC-ROs and positive score movement, or defection, which risks detection and negative score movement. The expected value of the honest choice depends on the fraction of other firms playing honestly, the long-run value of continued consortium membership, and the compounding effect of reputation over repeated interactions. The expected value of defection depends on the probability of detection, the severity of the consequences if detected, and the immediate gain from the specific dishonest action.

At the network scale envisaged by NCTE (modeled on a projected network of 700 firms, a scale chosen to illustrate consortium-grade dynamics), the detection probability P approaches 1 because the observation surface grows with the number of counterparties and the audit trail is immutable. The loss from detection includes exclusion from the consortium, which at scale becomes a catastrophic commercial cost. The gain from any single defection is bounded. The equilibrium therefore selects for honest behavior as the rational strategy. The rationality formula (P(detection) multiplied by Loss(reputation plus exclusion plus penalties) exceeds Gain(collusion)) is satisfied by wide margin at consortium scale.

This is a theoretical projection, not yet empirically validated. The 700-firm scale is illustrative. Validation at field scale is part of the research agenda in Section 8. The projected equilibrium is defensible, however, because it derives from well-established mechanism design principles, because analogous dynamics have been observed in existing voluntary coordination systems such as JARUS and SWIFT, and because the governance architecture is explicitly designed to produce the conditions under which the equilibrium becomes stable.

7.6 Governance and assurance implications

The governance architecture above has two implications worth making explicit. First, governance under NCTE is concentrated in the policy layer. Three additional governance artifacts sit alongside the five components above: a schema registry of standardized, versioned event object schemas; a role-object compatibility matrix specifying which roles may sign which object classes under which conditions; and deterministic signature policies defining thresholds, multi-signature requirements, and escalation paths. Changes to any of these artifacts are themselves PAC-ROs. The governance of the accounting system is governed by the accounting system. This recursion is how mature governance regimes operate.

Second, continuous auditing becomes feasible in a way it has not been under prior methods. The continuous assurance tradition [1, 2a, 28a] has been limited by the difficulty of constructing timely, reliable, coherently structured event data from heterogeneous enterprise systems. NEO Accounting improves feasibility by improving the auditable object. Auditors verify the integrity of PAC-ROs at event time rather than reconstructing these properties after the fact. Under NEO Accounting, a dispute is a PAC-RO of class Exception, linked to prior objects. Its resolution proceeds by inspection of the linked objects. Object inspection is faster, more reproducible, and less subject to reconstructive bias than narrative adjudication.

We emphasize what NEO Accounting does not claim. It does not claim to resolve disputes fairly in a substantive sense. A signature policy may itself be unjust. A schema may exclude relevant evidence. An exception taxonomy may fail to represent a marginalized grievance. NEO Accounting provides procedural accountability, the ability to detect, locate, and contest governance failures, which is a necessary but not sufficient condition for substantive fairness. The accountant's contribution is to produce clean records. Substantive justice requires work the accountant does not perform.

8. Research Agenda and Validation

NEO Accounting is a method claim. Method claims require empirical validation, and the research agenda is organized accordingly.

8.1 Horizon Europe Vanguard AI as the primary validation environment

The principal validation environment is the Horizon Europe Vanguard AI consortium [20], a sixteen-partner consortium across six EU Member States coordinated by 4D Consulting (Hungary) with partners including HashNET Technologies, COTRUGLI Business School, the University of Zagreb Faculty of Electrical Engineering and Computing, the Jožef Stefan Institute, Technische Universität Graz, Gdańsk University of Technology, Villanova.ai, Santer Reply, CloudFerro, and others spanning scientific leadership, sovereign infrastructure, cross-border federation, product engineering, and institutional validation.

Vanguard AI treats the Cotrugli Ledger as the constitutional governance layer for a multi-agent healthcare coordination system. The validation domain is professional well-being coordination across Croatia, Slovenia, and Hungary, with a prospective longitudinal study of at least 300 participants and at least 5,000 user-months of self-generated data. The project includes explicit GO/NO-GO gates at M12, M18, M24, and M30, and a staged methodology from architecture proving through

shadow-mode deployment to prospective validation. The methodology is transferable; the consortium includes a transferability benchmark based on 2,000 synthetic logistics and data pipeline scenarios.

The significance of the Horizon consortium for the claim of this paper is not that it provides empirical validation (that is future work). It is that a consortium of this size, scientific depth, and institutional diversity is prepared to treat NEO Accounting as the constitutional governance layer of a serious research program. Together with the formal method specification in Kapusta and Brčić [18] and the operational deployment on SI-Chain at national scale, this warrants the field-level claim of the paper.

8.2 Pilot domains beyond healthcare

Five additional domains offer high signal-to-noise ratios for validation of NEO Accounting as a general method: supply chain handoffs, regulated financial services, energy trading, high-value asset maintenance, and cross-border public-sector coordination. Each features costly cross-boundary truth, measurable dispute rates, and machine-speed transaction velocity.

8.3 Testable propositions

We treat five propositions as hypotheses under empirical test. None of them are asserted as findings. All of them are candidates for falsification.

Proposition 2 (Dispute reduction hypothesis).

We hypothesize that adoption of NEO Accounting reduces cross-boundary dispute frequency in the affected transaction classes. PBTE [18] pre-registers a threshold of at least 20% reduction relative to baseline matched classes. Whether this threshold is met is empirically open.

Proposition 3 (Resolution acceleration hypothesis).

We hypothesize that for disputes that do occur, NEO Accounting reduces time to resolution by shifting the resolution process from narrative reconstruction to object inspection. PBTE pre-registers a threshold of median exception resolution time at or below 75% of baseline.

Proposition 4 (Audit-effort reduction hypothesis).

We hypothesize that audit effort per unit of transaction volume decreases under NEO Accounting because evidentiary completeness is established at event time rather than constructed during audit.

Proposition 5 (Safe delegation hypothesis).

We hypothesize that NEO Accounting enables safer delegation of higher transaction volumes to AI agents under the IAAF framework, because agent authority is bounded by mandate credentials, actions are constrained by signature policies enforced at signing time, and kill switches tied to score thresholds produce automatic downgrade when integrity degrades.

Proposition 6 (Memory integrity hypothesis).

We hypothesize that when memory writes are governed by NEO Accounting, the structural record of each write makes corruption inspectable and retroactively detectable when evidence is later invalidated, producing a cleaner evidentiary baseline for memory integrity than unstructured writes alone. We make no claim that NEO Accounting prevents memory injection attacks or reduces their immediate

success rate. Whether combined deployment with retrieval-time defenses such as A-MemGuard produces additional benefit is an empirical question that requires collaborative testing with the security research community, which NEO Accounting supports by providing the provenance-anchored records against which such benefit can be measured.

Proposition 6 is the most speculative and the most valuable to test. It connects the accounting method framing directly to the open security problem and, if supported, would establish NEO Accounting as part of a deployable defense-in-depth configuration rather than only a governance framework.

9. The Coordination-Layer Pattern

Even with a sound method, a formal specification, a governance architecture, and a validation environment, the discipline becomes consequential only when the wider ecosystem adopts it. The question is not unique to NEO Accounting. It has been answered before, in other domains, by a pattern worth naming because it shapes how this discipline should be built, stewarded, and released.

When a new domain of human activity reaches a certain scale, every serious actor in that domain eventually needs a coordination layer, even when each actor would have preferred to dominate the domain alone. International civil aviation reached this point in 1944 and produced ICAO through the Chicago Convention. International banking reached it in 1973 and produced SWIFT. Container shipping reached it in 1968 and produced ISO 668. The internet reached it in the 1980s and produced DNS. Unmanned aviation reached it in 2007 and produced the Joint Authorities for Rulemaking on Unmanned Systems (JARUS), which now coordinates sixty-six national aviation authorities in the voluntary adoption of a shared rulebook for unmanned aircraft systems. The second author of this paper serves as Secretary General of JARUS and, in earlier work with Tronchetti [20a], developed the Exclusive Utilization Space concept as part of the conceptual substrate from which parts of that coordination framework emerged. The pattern across all five precedents is consistent. The coordination layer emerges not from top-down mandate but from bottom-up voluntary adoption by serious practitioners who recognize that the cost of bilateral coordination across borders exceeds the cost of submission to a shared framework.

Agent commerce is reaching its coordination layer moment now. Large platform operators need an interoperability layer because their agents will transact with agents they did not build. State actors need it because their agents will transact across jurisdictions whose legal frameworks do not currently recognize autonomous economic action. Financial institutions need it because settlement systems will be hit by agent-initiated volumes they cannot reconcile. Regulators need it because enforcement of emerging AI legislation requires auditable artifacts the current environment does not provide. None of these actors will welcome the standard a priori. All of them will converge on one eventually, because the alternative is unworkable. The question is not whether a coordination layer emerges. The question is whose architecture defines it and whether the architecture that defines it is honest. The Cotrugli Ledger is being built as a candidate architecture: model-neutral at the AI agent layer, vendor-agnostic at the DLT substrate layer, governed by an explicit registry hierarchy, grounded in a philosophical tradition that predates the

commercial interests of any single actor.

10. Limitations and Scope

We identify five limitations of the claim made in this paper.

First, the claim is a method claim, not a technology claim. NEO Accounting specifies what must be true of an accounting primitive for autonomous agents. It does not specify the implementation stack. Any distributed ledger substrate providing the required tamper-evidence, co-signature, credential binding, and timestamping guarantees may serve; the research program has used SI-Chain and HashNET, but Sui, MAPLE, SCITT, or custom substrates may equally serve.

Second, the claim is about accountability, not fairness. NEO Accounting makes governance failures detectable and contestable. It does not guarantee that governance will be substantively just. A signature policy may embed unjust authority relations. A schema may exclude relevant evidence. An exception taxonomy may fail to represent a marginalized grievance. Substantive fairness requires additional work not in scope here.

Third, the method presupposes credential infrastructure. In jurisdictions and ecosystems without mature credential infrastructure, NEO Accounting is not directly deployable. The EU Digital Identity Wallet ecosystem and W3C VC 2.0 provide the needed infrastructure in one major jurisdiction; other jurisdictions are at varying stages.

Fourth, the method presupposes enforceability. A PAC-RO is only as meaningful as the enforcement mechanisms that act on its validity. In commerce, enforcement is typically through contractual arrangements with counterparties and through downstream ledger postings. For agent memory, enforcement is through ingestion gates. Neither is automatic; both require deliberate deployment. The governance architecture of Section 7 is designed to make enforcement operationally tractable, but it depends on consortium formation, which is itself a coordination problem.

Fifth, the claim is historically situated. Accounting methods have evolved when the underlying economy changed. The claim that autonomous AI agents constitute a sufficient change is an empirical claim about AI deployment trajectory. If that trajectory stalls, the urgency of the claim diminishes. We regard the trajectory as robust given the 2025 convergence in both attack and defense work, but the urgency is conditional on it.

11. Conclusion

Accounting methods evolve when the economy changes. Pacioli's method was designed for a mercantile economy of human actors. Ijiri's extension was designed for an economy where rates of change matter. Grigg's extension was designed for an inter-party economy where cryptographic evidence could substitute for institutional trust. All three methods presuppose human actors. The emergence of autonomous AI agents as economic actors is a discontinuity of the same kind as the previous transitions. The conditions that made human-paced, narrative-reconstructive accounting methods adequate are absent for agent commerce. Something new is required.

AI needs at least a minimal truth layer, and one does not yet exist. The Cotrugli Ledger is the Truth and Evidence Layer designed to fill this gap. It is grounded in a civilizational tradition that runs from Ibn Khaldun's diagnosis of asabiyyah and institutional decay, through Cotrugli's articulation of honest records as the infrastructure that sustains long-distance cooperation, through Pacioli's mathematical codification, through the modern extensions of Ijiri and Grigg, to the formal framework and governance architecture specified in this paper and its companions. The method is narrow and honest in its claim: we do not prevent the world from going wrong, we produce the records under which it remains inspectable, contestable, and repairable when it does.

The discipline sits above a distributed ledger integrity substrate and below the existing enterprise accounting systems that consume its outputs. Its operational primitive is the Policy-Aware Co-signed Receipt Object. Its governance architecture is a five-component system of registries, twin scoring, dispute resolution, AI autonomy framework, and game-theoretic foundation. The research program is being validated at field scale through the Horizon Europe Vanguard AI consortium, on an operational substrate of SI-Chain powered by HashNET, at national scale. The discipline is model-neutral, vendor-agnostic, and multi-jurisdictional by design.

If accounting is civilization, and if civilization is extending to include autonomous artificial agents as economic actors, then upgrading accounting to accommodate them is a civilizational task. The Cotrugli Ledger is one coherent path to that upgrade. The rest is validation work, coordination work, and the long patient cultivation of an infrastructure whose success will ultimately be measured by how much of it becomes taken for granted.

Acknowledgements

This paper represents the convergence output of a three-year research program. The first author began this work in 2023 and developed the initial framing through discussions with colleagues at COTRUGLI Business School, HashNET Technologies, and the wider European sovereign AI and IPCEI-CIS/8ra ecosystem. The second author's work on global governance, JARUS coordination, and the Exclusive Utilization Space tradition shaped the coordination layer argument of Section 9 and the civilizational framing of Section 6. The third author's work on longitudinal reasoning and ambient intelligence at the Jožef Stefan Institute shaped the treatment of agent memory in Section 5 and the memory landslide diagnosis. The fourth author's work on explainable AI, multi-agent reasoning, and trustworthy memory architectures provided the technical grounding needed to formalize the method in Kapusta and Brčić [18] and informs the ongoing parallel research program on memory integrity. The fifth author's work on sovereign technical execution at HashNET Slovenia shaped the operational validation environment described in Section 1.3. The authors thank collaborators in the Horizon Europe Vanguard AI consortium for their commitment to treating the Cotrugli Ledger as the constitutional governance layer of a sixteen-partner research program. The authors also acknowledge constructive engagement with the authors of Wei et al. [30], Dong et al. [9], and the broader memory security research community, whose convergent work during 2025 has clarified the distinction between technical defense frameworks and method-level accounting discipline. Errors and opinions remain the authors' own.

Responsible AI Use Disclosure

The authors used multi-agent AI systems as assistive tools during manuscript preparation, including support for literature search, language refinement, formatting, and internal consistency checks. The research vision, problem framing, method design, argumentation, and final decisions are solely those of the authors. The authors reviewed, validated, and take full responsibility for all content in this paper.

Declarations

The first author is Principal and Founder of COTRUGLI Business School and Co-Founder and CEO of HashNET Technologies, which is developing implementations of the framework described in this paper. The second author is Secretary General of the Joint Authorities for Rulemaking on Unmanned Systems (JARUS), Executive Dean of the School of Global Governance at Beijing Institute of Technology, and Deputy Director of the National Research Center of ATM Law and Standard of China. The third author is Head of the Department of Intelligent Systems at the Jožef Stefan Institute. The fourth author is affiliated with the University of Zagreb Faculty of Electrical Engineering and Computing. The fifth author is affiliated with COTRUGLI Business School and serves in a technical execution capacity at HashNET Slovenia. The authors declare these interests. The framework itself is method-level and vendor-agnostic at the DLT substrate layer; implementations by any party, commercial or otherwise, are consistent with the research agenda proposed.

References

[1] Alles, M., Kogan, A., & Vasarhelyi, M. A. (2004). Putting continuous auditing theory into practice: Lessons from two pilot implementations. Journal of Information Systems, 18(s-1), 195–214.

[2] Cai, C. W. (2021). Triple-entry accounting with blockchain: How far have we come? Accounting & Finance, 61(1), 71–93.

[2a] Chan, D. Y., & Vasarhelyi, M. A. (2011). Innovation and practice of continuous auditing. International Journal of Accounting Information Systems, 12(2), 152–160.

[3] Cloudflare. (2025). Cloudflare and Coinbase launch the x402 Foundation to advance payments on the internet. Press release.

[4] Chen, Z., Xiang, Z., Xiao, C., Song, D., & Li, B. (2024). AgentPoison: Red-teaming LLM agents via poisoning memory or knowledge bases. Advances in Neural Information Processing Systems 37.

[5] Coinbase. (2025). x402: The internet-native payment protocol. Coinbase Developer Platform.

[6] Dai, J., & Vasarhelyi, M. A. (2017). Toward blockchain-based accounting and assurance. Journal of Information Systems, 31(3), 5–21.

[7] Cotrugli, B. (1573). Della mercatura e del mercante perfetto. Venice. [Manuscript completed in Naples, 1458.]

[8] Cotrugli, B. (2016). On Merchants and the Perfect Merchant. Published by COTRUGLI Business School and Dražen Kapusta. [First complete English translation.]

[9] Dong, S., Xu, S., He, P., Li, Y., Tang, J., Liu, T., Liu, H., & Xiang, Z. (2025). A practical memory injection attack against LLM agents. arXiv:2503.03704. Accepted NeurIPS 2025.

[10] Dončević, J., Fertalj, K., Brčić, M., & Kovač, M. (2024). Mask-Mediator-Wrapper architecture. IEEE Transactions on Software Engineering, 50(4), 900–910.

[10a] Ethereum Improvement Proposals. (2025). ERC-8004: Agent commerce standard. Draft.

[11] Gams, M., Luštrek, M., et al. (2019). Artificial Intelligence and Ambient Intelligence. Journal of Ambient Intelligence and Smart Environments, 11(1), 71–86.

[12] Google. (2025). AP2: Agent Payments Protocol. Technical specification.

[13] Grigg, I. (2005). Triple entry accounting. Working paper, iang.org. Also: Grigg, I. (2024). Triple entry accounting. Journal of Risk and Financial Management.

[14a] Ijiri, Y. (1982). Triple-entry bookkeeping and income momentum. American Accounting Association.

[14b] Ijiri, Y. (1986). A framework for triple-entry bookkeeping. The Accounting Review, 61(4), 745–759.

[14c] Ijiri, Y. (1989). Momentum accounting and triple-entry bookkeeping: Exploring the dynamic structure of accounting measurements. American Accounting Association.

[14d] European Union. (2024). Regulation (EU) 2024/1183 amending Regulation (EU) No 910/2014 as regards establishing the European Digital Identity Framework.

[15] Ibn Khaldun, A. (1377). The Muqaddimah: An Introduction to History. Translated by F. Rosenthal (1958). Princeton University Press.

[16] Krajna, A., Kovač, M., & Brčić, M. (2022). Explainable AI: An updated perspective. In Proceedings of MIPRO 2022, 859–864.

[17] Kapusta, D., & Liu, H. (2024). GENESIS: A proposal for a global coordination body for the transition from AI to AGI. Submitted to UNIDO, October 2024. On file with the United Nations Industrial Development Organization.

[18] Kapusta, D., & Brčić, M. (2026). Policy-bound triple-entry receipts for autonomous commerce. COTRUGLI Business School; University of Zagreb FER.

[19] Kapusta, D. (2026). NEO Cotruglian Triple Entry: A Truth and Evidence Layer for AI-era commerce. International Leadership Journal, forthcoming.

[20] Kapusta, D., et al. (2026). Vanguard AI: Next-generation AI agents with DLT-verified data provenance for personalized preventive healthcare. Horizon Europe RIA proposal submission.

[20a] Liu, H., & Tronchetti, F. (2019). The Exclusive Utilization Space: A new approach to the management and utilization of the near space. University of Pennsylvania Journal of International Law, 40(3), 555–594.

[21] Longo, L., Brčić, M., et al. (2024). Explainable AI (XAI) 2.0: A manifesto. Information Fusion, 106, 102301. Best Paper Award 2025.

[22] Machiavelli, N. (1513). Il Principe. Florence.

[23] Nau, D., Luštrek, M., Gams, M., et al. (2010). When is it better not to look ahead? Artificial Intelligence, 174, 1323–1338.

[24] MAPLE AI. (2026). The Agent Operating System: Architecture overview. Technical documentation.

[25] Pacioli, L. (1494). Summa de arithmetica, geometria, proportioni et proportionalità. Venice.

[26] Srivastava, S. S., & He, H. (2025). MemoryGraft: Persistent compromise of LLM agents via poisoned experience retrieval. arXiv:2512.16962.

[27] Sui Foundation. (2025). Verifiable AI control plane: Making AI accountable by design.

[28] Torra, V. (2026). Memory poisoning and secure multi-agent systems. arXiv:2603.20357.

[28a] Vasarhelyi, M. A., & Halper, F. B. (1991). The continuous audit of online systems. Auditing: A Journal of Practice & Theory, 10(1), 110–125.

[29] World Wide Web Consortium. (2025). Verifiable Credentials Data Model v2.0. W3C Recommendation, 15 May 2025.

[30] Wei, Q., Yang, T., Wang, Y., Li, X., Li, L., Yin, Z., Zhan, Y., Holz, T., Lin, Z., & Wang, X. (2025). A-MemGuard: A proactive defense framework for LLM-based agent memory. arXiv:2510.02373.

[30a] Kamimura, T. (2026). Verifiable AI Provenance Framework (VAP): An Architectural Framework for Evidentiary-Grade AI Decision Trails. Internet-Draft

draft-kamimura-vap-framework-00, VeritasChain Standards Organization, 8 January 2026. Available at https://datatracker.ietf.org/doc/draft-kamimura-vap-framework/.

[30b] Kamimura, T. (2025). SCITT Profile for Financial Trading Audit Trails: VeritasChain Protocol (VCP). Internet-Draft draft-kamimura-scitt-vcp-01, VeritasChain Standards Organization, 17 December 2025. Available at https://datatracker.ietf.org/doc/draft-kamimura-scitt-vcp/.