GBBC
Global Blockchain
Business Council

STANDALONE REPORT

AGENTIC AI & BLOCKCHAIN CONVERGENCE

GLOBAL STANDARDS MAPPING INITIATIVE (GSMI) 7.0

GBBC GSMI 7.0

GLOBAL BLOCKCHAIN BUSINESS COUNCIL

DC Location:
1629 K St. NW, Suite 300
Washington, DC 20006

Geneva Location:
Rue de Lyon 42B
1203 Geneva
Switzerland

THANK YOU TO OUR CO-CHAIRS

Dino Dell'Accio*
United Nations Joint Staff Pension Fund (UNJSPF)

Paul Rapino
Hashgraph

**Dino Dell'Accio is participating in his personal capacity and not as a representative of the United Nations.*

A full list of GSMI contributing organizations can be found on page 43.

TABLE OF

CONTENTS

SECTION I

INTRODUCTION

Artificial intelligence is entering a new phase in which systems are increasingly capable not only of generating information, but also of taking action. Agentic AI represents an important development in this evolution. Rather than responding only to individual prompts, agentic systems can pursue goals through multi-step processes that involve planning, reasoning, interacting with external tools and systems, making decisions, and executing actions with varying degrees of human involvement. Their autonomy can range from tightly prescribed workflows and human-in-the-loop validation to more dynamic execution within defined permissions and constraints. In this sense, agentic AI marks a progression from AI primarily used to generate or analyze information toward action-oriented applications capable of participating directly in economic and operational activity.

This evolution also makes questions of trust, identity, authority, validation, accountability, and governance increasingly consequential. An AI-generated recommendation can be reviewed before a human acts upon it; an autonomous agent, by contrast, may itself initiate a payment, interact with another agent, access an enterprise system, invoke a software tool, or execute a transaction. As the distance between AI decision-making and real-world action narrows, organizations need mechanisms to establish who or what an agent is, on whose behalf it is acting, what authority has been delegated to it, which rules govern its behavior, and how its actions can subsequently be verified.

These questions become still more complex in multi-agent environments in which actions and responsibility may extend across organizations, platforms, and jurisdictions.

It is within this context that agentic AI and blockchain technology can be understood as complementary technologies. Agentic AI introduces intelligence, adaptability, and autonomous execution; blockchain can support identity, provenance, coordination, programmable controls, and tamper-evident records. Blockchain does not make an AI model inherently accurate, unbiased, explainable, or secure, nor does it eliminate the need for human governance and conventional AI controls.

Rather, it can provide additional trust infrastructure through which selected elements of an agent's lifecycle—from identity and authorization to transactions and other material actions—can be registered, validated, and audited where shared verification is useful. The distinction between probabilistic AI behavior and the rule-based execution of connected transactions is central to understanding where their convergence may provide practical value.

This report builds on GBBC's previous work examining the convergence of AI and blockchain through the Global Standards Mapping Initiative (GSMI; see **Table 1**).

TABLE 1: GSMI'S AI-FOCUSED REPORTS

Report Version	Core AI Convergence Focus	Key Governance Milestone
GSMI 4.0[1]	Standalone updates detailing early integrations and foundational enterprise concepts	Initial regulatory alignments and defining baseline interoperability for data provenance
GSMI 5.0[2]	Strategic enterprise use cases and optimization of foundation models	Establishing key principles for AI growth and mapping early foundations to blockchain governance
GSMI 6.0[3]	Unpacking open-source and decentralized AI toward realistic, transparent, and reliable models	Securing event-level auditability and establishing blockchain as a governance layer for AI

For GSMI 7.0, this report advances GBBC's earlier work on the convergence of AI and blockchain. It focuses on the practical implications that emerge as AI moves from generating outputs to autonomously taking actions. It examines:

- **Definitions and distinguishing characteristics of agentic AI**
- **How blockchain capabilities may complement agentic systems**
- **Existing and emerging use cases**
- **Risks and mitigation approaches**
- **Governance, regulatory, and standards considerations**
- **Principles and recommendations for trustworthy implementation**

GSMI 7.0 therefore shifts the primary unit of analysis from an AI output to an autonomous action. Its organizing framework follows the accountability chain from principal, to agent identity, mandate and limits, through to the execution record, institutional responsibility, and contestability and redress.

Autonomy may change how an action is performed, but it should not sever any link in that chain. Section VI develops this accountability framework and applies it across the agent lifecycle.

The use-case analysis seeks to move beyond conceptual possibilities toward areas of meaningful adoption. It examines the particular role performed by agentic AI, the function blockchain provides, relevant real-world developments, and the limitations of the available evidence. Openness and decentralization do not in themselves guarantee trustworthy outcomes. Agentic systems still require appropriate governance, security, identity, authorization, accountability, and risk controls, while the supporting infrastructure must be assessed in light of the context and risks of each application. The report does not prescribe a single centralized or decentralized architecture; rather, it considers how trustworthy infrastructure can support a wider range of organizations, developers, and users.

Blockchain is most useful to agentic AI not as a source of intelligence or truth, but as infrastructure for identity, delegated authority, defined policy controls, shared evidence and accountability. Its use is justified where participants need shared verification, programmable controls over connected transactions, or evidence that is not held solely by one operator. Where those conditions are absent, it may add complexity without adding commensurate assurance.

SECTION II

WHAT IS AGENTIC AI?

Agentic AI is best understood as a way of operating rather than a settled legal category. It describes systems that can pursue an objective across several steps—selecting tools, drawing on external data and adjusting to new information within the permissions they have been given.

At its core, an AI agent is designed to pursue multi-step goals with varying degrees of autonomy, ranging from human-in-the-loop validation to more autonomous execution. For example, an agentic AI assistant could receive the instruction to organize a business trip and then, with appropriate permissions, search for flights, compare hotels, coordinate calendars, draft emails, make bookings, and monitor itinerary changes. Rather than executing a single task, the agent can manage an end-to-end workflow while continuously evaluating progress toward the user's objective. In enterprise settings, the same pattern can automate complex, knowledge-intensive processes.

Agentic AI is not a separate legal category under the EU AI Act.[4] An agentic system may nevertheless fall within the Act's technology-neutral definition of an AI system, with obligations depending on the actor's role, the system's intended purpose and the applicable risk classification. An agentic architecture does not, by itself, make a system high-risk. In the United States, the National Artificial Intelligence Initiative Act likewise defines AI broadly and does not create a separate category for agentic systems.[5]

There are, however, a number of existing definitions and interpretations among major stakeholders, grouped below as regulators and government bodies, formally recognized standards setters, industry, and academia.

REGULATORS & GOVERNMENT BODIES

EU AI ACT[6]

"AI system" means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.

U.S. NATIONAL ARTIFICIAL INTELLIGENCE INITIATIVE ACT OF 2020 (15 U.S.C. § 9401)[7]

The term "artificial intelligence" means a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments. Artificial intelligence systems use machine and human-based inputs to—

(A) perceive real and virtual environments;
(B) abstract such perceptions into models through analysis in an automated manner; and
(C) use model inference to formulate options for information or action.

SINGAPORE'S MODEL AI GOVERNANCE FRAMEWORK FOR AGENTIC AI (VERSION 1.5, 2026)[8]

Agentic AI systems are software systems consisting of one or multiple AI agents that may operate individually or collaboratively (working definition).

FORMALLY RECOGNIZED STANDARDS SETTERS

ISO/IEC 22989[9]

An AI agent is an automated entity that senses and responds to its environment and takes actions to achieve its goals.

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (NIST)[10]

AI agents are software programs that can interact with their environment, receive information, and undertake self-directed actions in service of a larger, externally-specified goal.

IEEE[11]

An intelligent agent is a computational system situated in an environment that is capable of flexible, autonomous action to achieve design objectives.

ITU FOCUS GROUP ON TRUST AND IDENTITY FOR HUMANS AND AGENTIC AI (2026)[12, 13]

An Agentic AI or AI Agent or Agent is an autonomous software entity that interprets goals, formulates intent, and executes actions, often by invoking external tools, APIs, or other agents to achieve outcomes on behalf of users. Its behavior is driven by model-based reasoning, typically relying on Large Language Models (LLMs) or other Foundation Models (FMs) to analyze context, generate plans, and adapt its actions dynamically.

INDUSTRY

ANTHROPIC[14]

"Agent" can be defined in several ways. Some customers define agents as fully autonomous systems that operate independently over extended periods, using various tools to accomplish complex tasks. Others use the term to describe more prescriptive implementations that follow predefined workflows. At Anthropic, we categorize all these variations as agentic systems, but draw an important architectural distinction between workflows and agents:

- Workflows are systems where LLMs and tools are orchestrated through predefined code paths.
- Agents, on the other hand, are systems where LLMs dynamically direct their own processes and tool usage, maintaining control over how they accomplish tasks.

IBM[15]

An artificial intelligence (AI) agent is a system that autonomously performs tasks by designing workflows with available tools.

GOOGLE CLOUD[16]

AI agents are software systems that use AI to pursue goals and complete tasks on behalf of users. They show reasoning, planning and memory and have a level of autonomy to make decisions, learn and adapt.

Their capabilities are made possible in large part by the multimodal capacity of generative AI and AI foundation models. AI agents can process multimodal information like text, voice, video, audio, code, and more simultaneously; can converse, reason, learn, and make decisions. They can learn over time and facilitate transactions and business processes. Agents can work with other agents to coordinate and perform more complex workflows.

ACADEMIA

MIT MEDIA LAB[17]

Agentic AI refers to AI systems designed to act autonomously, perceiving their environment, making decisions, and taking actions to achieve specific goals. These systems often incorporate features like adaptability, goal orientation, and interaction with dynamic environments.

STANFORD INSTITUTE FOR HUMAN-CENTERED ARTIFICIAL INTELLIGENCE (HAI)[18]

Agentic AI refers to AI systems designed to act as autonomous or semi-autonomous agents: they can set or interpret goals, plan and sequence actions, use tools (like web browsers, code, or APIs), make decisions based on feedback, and adapt over time to complete tasks. Unlike a purely reactive chatbot that only responds turn-by-turn, agentic AI is oriented around ongoing task execution—breaking down objectives, coordinating steps, and sometimes operating with minimal human oversight within defined constraints.

SECTION III

HOW AGENTIC AI & BLOCKCHAIN GO TOGETHER

Agentic AI and blockchain can complement one another, but the value of combining them depends on the workflow. An agentic system may select and carry out actions; blockchain or another form of distributed ledger technology (DLT) may give several participants a shared, tamper-evident record and, when the workflow uses smart contracts, enforce defined transaction rules. Whether that combination is useful turns on what is recorded, how off-chain events are authenticated, who governs the network and whether the participants genuinely need a shared means of verification.

Three roles are involved here, and they should not be conflated. Enforcement determines whether an action is blocked, conditioned or routed; an evidence layer records selected facts about what happened; and governance assigns authority, oversight, responsibility and remedy. A system can perform one of these roles without performing the others.

The IMF's 2026 note on agentic AI in payments[19] offers a useful way to think about this separation. Its three-layer model places probabilistic reasoning in intent and orchestration, while control and authorization remain mandate-based and deterministic, and settlement retains legal finality. Keeping those layers distinct helps prevent an agent's adaptive reasoning from becoming an unchecked payment instruction.

Blockchain can preserve a tamper-evident record of transactions carried out by AI agents, together with cryptographic commitments to selected off-chain evidence such as inputs, model and policy versions, and approvals. That record shows what was committed to the ledger, but not whether the underlying evidence was accurate, complete or lawful; those questions still depend on how the evidence was produced and governed. Blockchain can also support asset ownership and transfer, programmable payments, interaction with smart contracts, and the verification of identities or credentials. Where the surrounding workflow has been designed accordingly, smart contracts can enforce defined business rules and transaction limits.

This combination may be useful in enterprise, financial-services and public-sector settings where several parties need a shared record of agreed events. An AI agent might, for example, negotiate a service agreement within a defined mandate, verify a counterparty's digital credentials, make a payment using an approved instrument and record the transaction in a tamper-evident audit trail. Whether an external party can verify that trail without relying on the operator will depend on how the evidence is generated, controlled and anchored.

One vendor-reported pilot, the Verifiable Governance and Sovereignty initiative from Hedera Foundation, EQTY Lab and Accenture, combines runtime cryptographic evidence of AI behavior with tamper-evident logging for proposed use in government environments. It illustrates the approach but does not establish wider public-sector adoption.

For financial services and digital assets, agents could automate workflows across various activities, including tasks such as monitoring market conditions, executing programmable payments, performing compliance checks, reconciling transactions, managing collateral, optimizing treasury operations, or coordinating post-trade activities across multiple institutions.

Combining agentic AI with blockchain also creates governance questions of its own. AI behavior is probabilistic, while smart contracts can execute predefined rules deterministically for the same state and inputs. The resulting record and settlement remain subject to the network's consensus and finality model, as well as oracle inputs, forks, reorganizations and cross-chain dependencies. DLT cannot contain arbitrary off-chain conduct, correct a flawed objective or remove bias from human inputs. It may preserve the provenance of data or attestations supplied to an agent, but it cannot establish that the underlying material is accurate, complete, representative or fair.

Whatever architecture is chosen, organizations will still need clear policies for agent identity and authorization, accountability, cybersecurity, privacy and risk management. Digital identity, verifiable credentials, programmable controls and auditable smart contracts can support those policies where they fit the use case, but they are not substitutes for them.

USE CASES OF AGENTIC AI AND BLOCKCHAIN

Table 2 on page 14 brings together examples of agentic AI and blockchain being tested or deployed, alongside several adjacent AI and blockchain applications retained from earlier GSMI reports. For each, it describes the relevant AI or agentic function, the role that blockchain or DLT may play, the apparent maturity of the cited development and the main limitation in the available evidence. Inclusion should not be read as a claim that blockchain is necessary for the use case or that one cited project represents wider adoption.

These arrangements work only if their incentives and governance are sound and the data they rely on are trustworthy. Blockchain may help distribute incentives or preserve records, but recording an AI decision does not align participants' interests or validate the quality of that decision.

TABLE 2: ILLUSTRATIVE DEVELOPMENTS AND USE CASES

The maturity rating applies only to the cited example, not to the use-case category as a whole. The scale runs from Research/concept, through Prototype and Pilot, to Limited production and Scaled production. Ratings are based on public information reviewed through September 9, 2026, and vendor-reported deployments have not been independently verified. Several adjacent AI and blockchain applications from earlier GSMI reports are retained and clearly marked so that they are not mistaken for agentic deployments.

Use Case / Application	Agent and DLT Functions	Illustrative Source or Implementation
Autonomous AI Agents *Maturity (indicative): Research/concept*	Within a defined mandate, an autonomous agent may select and carry out actions. Smart contracts and decentralized governance mechanisms could publish policy parameters, record approved changes and enforce rules for transactions routed on-chain. Voting, however, does not ensure that those rules are appropriate, and it cannot govern conduct that remains off-chain.	ETHOS Framework (research paper)
AI-Enhanced Decentralized Physical Infrastructure (DePIN) (Mentioned in GSMI 6.0) *Maturity (indicative): Research/concept*	Here, agents could forecast demand and allocate compute, sensor or storage capacity, while DLT coordinates payments and incentives and records the resources provided across the network. The cited material describes an emerging design pattern rather than evidence of a scaled autonomous deployment.	Supra overview of AI and DePIN
Graph AI for Blockchain AML (Adjacent Application) *Maturity (indicative): Prototype*	The cited prototype applies graph models to blockchain transaction data, generating anomaly signals and explanations for human review. It does not describe an autonomous agent or the use of smart contracts to trigger compliance action.	Graph-based crypto anomaly detection (research paper)
Zero-Knowledge Biometric Authentication (Adjacent Application; addressed in GSMI 5.0 and GSMI 6.0) *Maturity (indicative): Prototype*	An agent could request or present a credential claim, although the cited source concerns identity technology rather than an agentic deployment. Decentralized identifiers and credential-status registries may support verification, while zero-knowledge proofs can disclose a specific claim without revealing the underlying biometric data. Verification itself remains off-chain and depends on issuer, liveness and key-management controls.	Ping Identity overview of zero-knowledge biometrics

Use Case / Application	Agent and DLT Functions	Illustrative Source or Implementation
Agentic Commerce *Maturity (indicative): Pilot*	An agent may discover products, compare offers, negotiate within a mandate and assemble a purchase instruction. DLT can support wallet identity, programmable payment conditions and a transaction record shared across the parties. The agent's probabilistic reasoning should remain separate from deterministic authorization and legally final settlement; stablecoin transactions must also account for the rules applying to the instrument and issuer, as well as effective redress.[20]	Visa Intelligent Commerce; x402 protocol
Public Services *Maturity (indicative): Pilot*	Agents may operate within defined emergency-management or cyber-response workflows, with DLT recording timestamped commitments to selected runtime evidence and policy events. The cited case is a vendor-reported pilot rather than evidence of wider government adoption, and any deployment would still require clear authority, human review, procurement controls and lawful use of data.	Hedera Foundation, EQTY Lab and Accenture case study (vendor-reported pilot)
AI Tokens (Adjacent Ecosystem Example)[21] *Maturity (indicative): Scaled production for the cited example*	Numerai crowdsources predictive models, making it an adjacent AI and blockchain example rather than evidence of agentic AI. Its cryptoasset helps coordinate incentives, access or governance within the network. As with any token, the applicable legal treatment depends on its rights, functions and use.	Numerai
Business-to-Consumer Operating Models *Maturity (indicative): Research/concept*	Consumer and enterprise agents may exchange information or transact across organizational boundaries. Verifiable credentials, permissions, consent records and shared event evidence could reduce the need for either party to rely solely on the other's systems. The cited source does not establish a concrete deployment, however, and any practical model would need to address control of consumer data, conflicts of interest and redress.	SingularityNET (industry example)
Agent-to-Agent Communications *Maturity (indicative): Limited production*	Agents may need to authenticate one another, exchange information, negotiate terms and coordinate tasks. Where participants require a common basis for verification, DLT can provide shared credential status, delegation records or transaction evidence. Open standards for identity, messaging and payments may provide interoperability without a ledger.	Fetch.ai and Artificial Superintelligence Alliance (industry examples)
Verifiable AI Supply Chain *Maturity (indicative): Pilot*	In a controlled workflow, the agent's software, policy and runtime state can be identified, while DLT preserves consensus-ordered, timestamped commitments to selected prompts, code, model versions and runtime events. These records strengthen traceability, but they do not show that the underlying material is complete, lawful or fit for purpose.	EQTY Lab AI Integrity Suite; Hashgraph, EQTY Lab, and Accenture pilot (vendor-reported)
Mandate and Guardrail Evidence *Maturity (indicative): Research/concept*	Before an agent undertakes a delegated task, DLT can preserve a signed, timestamped commitment to the applicable policy version, permission scope, limits and approvals. The resulting record shows what was specified and when; it does not establish that the guardrails were adequate, that every event was captured or how liability should be allocated.	Artzt and Fei, Solicitors Journal

TABLE 3: EMERGING USE CASES

Concept	Indicative Status	Key Advantage
Zero-Knowledge Machine Learning (ZKML)	Emerging technique	Can cryptographically verify that a specified computation was executed on committed inputs using a committed model, without revealing protected inputs. It does not establish model quality or the substantive correctness of the outcome; applicability and cost depend on the implementation.
Fully Homomorphic Encryption for AI and Blockchain	Emerging technique	Allows computation on encrypted data; performance, key management and implementation security remain material constraints.
Agentic AI Wallets	Emerging implementation pattern	May enable agents to hold credentials and initiate approved on-chain transactions within wallet and policy limits.
Decentralized AI Compute (DePIN)	Emerging implementation pattern	May support more distributed access to AI compute infrastructure; service quality and concentration still require assessment.
Data Sovereignty Protocols	Concept	May support consent, access control and compensation arrangements for data use.
Quantum-Ready Protocols	Research/concept	Supports migration to cryptography designed to resist quantum attacks; standardization and migration are still evolving, and implementation remains at an early stage.
Artificial Superintelligence Alliance (ASI) Decentralized AI	Industry platform example	Provides alternative infrastructure for decentralized AI services.
Verifiable Content Authentication	Operational provenance tools; agentic use emerging	Tamper-evident provenance records can support content authentication; they do not establish that content is accurate or truthful.
Know Your Agent	Specification in development	Verifiable credentials and signed delegations can help counterparties check an agent's identity, principal, mandate, permission scope and status.
External Verification of Agent Control Adherence	Standards work in progress	May allow relying parties to verify that an agent acted within defined controls without depending solely on records held by the operator.

BENEFITS OF AGENTIC AI

In the context of blockchain and digital assets, agentic AI can go beyond simply assisting users. AI agents can execute business processes, coordinate across organizations, and respond dynamically to changing conditions. Across the existing and emerging use cases described above, potential benefits include:

- **Automation of complex workflows:** Agents can coordinate multi-step processes such as reconciliation, collateral management, treasury operations and compliance monitoring. The gains will depend on whether the process is suitable for automation, the data are reliable and exceptions can be handled effectively.
- **Faster decision-making:** By testing current data against defined policies, agents can propose or initiate action more quickly than a manual workflow. Where the decision carries significant consequences, suitable limits, validation and escalation should remain in place.
- **Execution of programmable financial operations:** Where an agent has the necessary authorization, it can initiate a transaction involving a smart contract, stablecoin or tokenized asset once defined conditions are met. The payment authorization layer should apply predefined rules, while the settlement arrangement should make the network's finality model and the point of legal finality explicit.
- **Improved compliance and risk management:** Agents can monitor transactions and workflows against defined indicators for fraud, sanctions, AML/CFT, operational risk and internal policy, then create a record for review. Where several parties need the same tamper-evident evidence, blockchain can preserve selected events.
- **Greater interoperability and coordination:** Agents may coordinate work across networks, institutions and enterprise systems, but successful interoperability still rests on compatible standards for identity, messaging, data and payments.
- **24/7 operations and scalability:** Agents can monitor systems and handle routine activity continuously, which may be valuable in always-on markets. This requires escalation and intervention arrangements that work outside ordinary business hours.
- **Verifiable delegation and control adherence:** Organizations have a stronger basis for delegating authority when material agent actions leave tamper-evident evidence of the applicable mandate, approvals and control checks. If parties outside the operator's systems can verify that evidence, it may also support assurance across organizational boundaries.

SECTION IV

RISKS OF AGENTIC AI & MITIGATION APPROACHES

When an agent acts across several accounts, tools or organizations, attributing an action can become difficult. The practical questions are who authorized it, what mandate and limits applied, what evidence was retained and which institution remains responsible. Model opacity creates a separate problem for explanation and validation, but it does not remove the need to trace authority and reconstruct execution.

Some risks arise within a single organization's agent environment; others emerge when agents interact across organizational boundaries. The distinction matters because attribution, control and coordination become progressively harder as more systems and institutions are involved. Multi-agent risk also needs to be considered at ecosystem level. Google DeepMind's June 2026 research call points to the possibility that interacting populations of agents could display capabilities or failure modes that are not apparent when each model is tested in isolation.[22] The scenarios include miscoordination, conflict, collusion, destabilizing feedback and new security vulnerabilities. This creates a need for shared testbeds, better understanding of agent networks, identity and reputation infrastructure, and oversight at population level.

Agentic systems bring familiar operational, technical, legal and governance risks into a less familiar setting. The table therefore separates the conventional controls already available from the vulnerabilities that remain, and identifies both potential DLT functions and the non-DLT controls that are still required. DLT may help preserve evidence or enforce rules for connected transactions, but it is not a general answer to AI risk.

For practical purposes, the table groups the risks under governance and accountability; legal and compliance; model and data; identity and access management; cybersecurity; technology and infrastructure; privacy; records and assurance; third-party dependencies; and systemic or market risk. These categories inevitably overlap and should be read as a control map, not as a rigid taxonomy.

Annex 1 on page 40 applies a similar set of categories to a contributor-supplied view of the OpenAI and ChatGPT ecosystem. It is intended as an illustrative control map, not as a verified inventory of OpenAI's controls.

TABLE 4: AI AGENT RISKS AND CONTROLS

Risk Scenario or Failure Mode	Existing AI Controls	Residual Vulnerabilities	Potential DLT Contribution	Complementary Control or Limitation
Autonomous decision-making errors: This includes unauthorized or out-of-bounds autonomous action, such as sandbox escape. AI agents may misinterpret objectives, rely on inaccurate data, or produce unintended outcomes, potentially resulting in erroneous transactions, financial losses, or operational disruptions.	Sandboxing; tool allowlists; least-privilege permissions; approval gates; runtime monitoring; circuit breakers; pre-deployment capability and safety evaluations; and documented system limitations.	An agent may chain vulnerabilities, leave the intended control boundary, use compromised credentials, or act before detection.	• Register agent and credential identity • Attest mandate and policy versions • Record signed action checkpoints • Support credential revocation • Enforce transaction limits where execution passes through an integrated smart contract	Blockchain does not contain a model or stop arbitrary off-chain conduct. Secure infrastructure, continuous detection, incident response and legal responsibility remain necessary.
Cybersecurity threats: AI agents may become targets for hacking, prompt injection, data poisoning, credential theft, or exploitation of vulnerabilities in APIs, wallets, or smart contracts, potentially compromising digital assets or sensitive information.	Prompt-injection testing and filtering; credential isolation and rotation; secure tool and API integration; zero-trust access; adversarial testing; security monitoring; and incident response. Relevant reference points include provider secure-AI frameworks and the OWASP guidance for agentic applications.	Prompt injection, compromised credentials, malicious tools, and poisoned inputs may remain effective even where later events are recorded.	Register agent and credential status; preserve signed security events and code or policy versions; support rapid credential revocation and transaction limits where execution is integrated.	DLT does not secure models, endpoints, keys, APIs, or off-chain infrastructure. Secure development, zero-trust access, monitoring, and incident response remain essential.
Governance and accountability: It may be difficult to determine responsibility when an autonomous agent makes an incorrect decision or executes an unauthorized transaction. Clear governance frameworks, approval thresholds, and human oversight remain essential.	Defined accountable owners; risk-tiered approvals; model and system documentation; independent review for higher-risk uses; change control; escalation routes; incident reporting; and board or senior-management oversight proportionate to impact.	Human approval may degrade as volume rises; reviewers may lack sufficient context or receive a selectively framed request; and a recorded approval does not show that it was adequately considered.	Commit to the request as presented, the applicable policy version, the approval outcome, the approver's identity, and the timestamp as a linked evidence set.	A verifiable approval record does not establish the quality of the review. Staffing, workload, escalation design, and oversight thresholds remain organizational responsibilities.

Risk Scenario or Failure Mode	Existing AI Controls	Residual Vulnerabilities	Potential DLT Contribution	Complementary Control or Limitation
Compliance and regulatory risk: AI agents must operate in accordance with evolving regulatory requirements related to AML/CFT, sanctions, privacy, consumer protection, securities laws, and recordkeeping. Failure to comply could expose organizations to legal and regulatory consequences.	Policy mapping; AML/CFT and sanctions controls where relevant; privacy and consumer-protection review; jurisdictional checks; recordkeeping; human review of exceptions; and periodic testing against changing legal requirements.	Legal duties and policy interpretations change; off-chain facts may be incomplete; and automated controls may not account for jurisdiction, consumer redress, or exceptional circumstances.	Record policy versions, approvals, attestations, and transaction events; apply programmable controls and selective disclosure where appropriate.	DLT does not determine legal interpretation, establish the truth of KYC or screening data, or replace human review, supervisory judgment, and effective redress.
Identity and authorization risks: This includes identity, authority or attribution failure in a multi-agent chain. AI agents require trusted digital identities and carefully managed permissions. Weak authentication or excessive privileges could enable unauthorized actions or fraudulent transactions.	Enterprise identity and access management; SSO and role-based access controls; scoped service accounts and API credentials; signed authorization records; least privilege; separation of duties; credential monitoring; rotation; and rapid revocation.	• Keys may be stolen • Authority may be over-broad • Agents may create sub-agents • Causation may cross organizations and systems • Permissions, tools, and integrations may also accumulate without triggering reassessment, causing the deployed agent to diverge from the assessed configuration.	• Use verifiable agent credentials based on open standards. For example, KYA-OS (Know Your Agent Operating System, a DIF specification) uses Decentralized Identifiers (DIDs) for agent identity and W3C Verifiable Credentials to represent scoped, revocable delegation • Record delegation and sub-delegation, and assess the complete chain against the principal's original mandate; verifying each step does not by itself establish that the composite delegation remains in scope • Maintain principal-agent relationships • Publish status and revocation • Preserve shared evidence across participants • Maintain a signed, versioned record of the agent's authorized envelope and subsequent changes.	Legal attribution and responsibility still depend on law, contract, institutional governance, and the facts of the case. Versioning shows when authority changed; it does not determine whether the resulting scope remains appropriate.

Risk Scenario or Failure Mode	Existing AI Controls	Residual Vulnerabilities	Potential DLT Contribution	Complementary Control or Limitation
Smart contract and infrastructure vulnerabilities: AI agents interacting with smart contracts inherit the risks of software bugs, coding errors, oracle failures, blockchain outages, and interoperability issues across networks.	Software and smart-contract review; independent audit; formal verification where proportionate; model and software version management; oracle and bridge controls; staged deployment; monitoring; business continuity; and recovery procedures.	Deterministic systems may execute flawed code or inaccurate oracle data; bridges, networks, and interfaces may fail; and transactions may be difficult to reverse.	Attest code and policy versions; record governance changes; use multisignature approvals, timelocks, transaction caps, and circuit breakers where supported.	Independent audits, testing, oracle resilience, key management, network monitoring, and off-chain recovery arrangements remain necessary.
Privacy and confidentiality concerns: AI agents often require access to sensitive financial and customer data. Organizations must ensure appropriate data governance, privacy protections, and confidential computing techniques where appropriate.	Data minimization; purpose limitation; enterprise privacy and retention controls; redaction and tokenization; encryption; confidential-computing techniques where appropriate; access logging; and privacy impact assessment.	Persistent records and metadata may conflict with data minimization, correction, confidentiality, and deletion requirements.	Keep sensitive data off-chain; use hashes, verifiable credentials, zero-knowledge proofs, selective disclosure, and credential-status mechanisms where appropriate.	Privacy impact assessment, lawful basis, access controls, key management, and correction or deletion processes remain necessary. A ledger should not hold raw personal or confidential data by default.
Model reliability and bias: This includes erroneous, biased or manipulated decisions caused by data or model failures. AI systems may generate inconsistent or difficult-to-explain decisions, complicating validation, audit and regulatory review.	Representative test data; model evaluation and red-teaming; bias and robustness testing; retrieval grounding; structured outputs; version control; drift monitoring; human review for consequential decisions; and published limitations.	Data provenance may be uncertain; models and policies may drift; poisoned inputs may appear valid; and reasoning may remain opaque.	• Anchor dataset, model and policy hashes • Attest source, version and time • Maintain a tamper-evident provenance and change record	A valid record does not prove that data are true, representative or fair. Independent validation and impact assessment remain necessary.
Third-party and concentration risk: Agents may depend on common model providers, cloud platforms, APIs, identity services, data sources, oracles and bridges. An outage, compromise, policy change or correlated model behavior can disrupt several organizations.	Dependency inventory; concentration limits; vendor due diligence; portability testing; contractual service and incident terms; fallback providers or manual procedures; exit planning; and monitoring of external services.	Alternative providers may not be equivalent; switching can be slow; dependencies may be shared or opaque; model behavior can change; and fallback procedures may not operate at machine speed.	Shared service-status or revocation records; portable credential registries; multiple data sources or oracles where justified; and tamper-evident service attestations.	Distributed records do not remove dependence on models, cloud infrastructure, keys, data or legal contracts. Resilient alternatives, governance and supervisory powers remain necessary.

Risk Scenario or Failure Mode	Existing AI Controls	Residual Vulnerabilities	Potential DLT Contribution	Complementary Control or Limitation
Systemic and market risk: Widespread use of similar models or agent strategies may create correlated behavior, feedback loops or rapid propagation of errors across institutions and markets.	Phased rollout; transaction and rate limits; real-time telemetry; kill switches and graded intervention; concentration and third-party risk management; scenario testing; cross-firm incident coordination; and regulatory monitoring of correlated behavior.	• Machine-speed actions may cascade • Cross-platform effects may be hard to halt • Payments may be irreversible • Jurisdiction and remedy may be unclear	Apply smart-contract guardrails, multisignature approval, escrow, transaction caps, shared event records and coordinated revocation.	On-chain controls govern only actions and assets connected to them. Off-chain recovery, consumer protection, insolvency rules and regulatory coordination remain necessary.
Evidence integrity and repudiation: Logs may be generated and controlled by the system or operator under examination, allowing material events to be omitted, altered, or selectively disclosed.	Application and agent logging; security information and event management (SIEM) ingestion; defined retention; append-only or write-once storage; separation of operating and logging functions; independent review of logging controls.	The operator may still determine what is recorded and disclosed; retention may expire; and an audit of logging controls does not prove that a particular event record is complete.	Time-bound cryptographic commitments; digitally signed evidence binding actions to authenticated identities and mandates; timestamps; external anchoring; selective disclosure.	Anchoring can show that a disclosed record matches an earlier commitment and has not changed since then. It does not establish completeness, truth, fairness or lawfulness. Even where execution requires evidence generation, completeness still depends on whether the control can be bypassed and whether relevant off-chain events are captured.

RISK MITIGATION

Governance should cover an agent's full lifecycle, from design and deployment through day-to-day operation and eventual retirement. Organizations need to define the agent's mandate and permissions, apply least privilege and risk-based human approval, validate the models, data and tools it uses, and monitor its activity in operation. They also need workable procedures for preserving audit evidence, responding to incidents, revoking access and providing redress. Where several parties must rely on the same credentials, policy attestations or event records, blockchain may complement these arrangements, but it cannot replace model assurance, cybersecurity, legal responsibility or human oversight.

Trustworthy agentic systems depend on a traceable line of provenance. That includes the lineage of the data, the model, algorithm, policy and software versions in use, the tools called and authorizations given, and a cryptographic chain of custody for material records.

The practical test is not simply whether a record is immutable, but whether the evidence leads back to sources that are reliable enough for the use case and its risks. Because agents may make decisions or initiate transactions with limited human involvement, policy should state clearly what they may do and when they must stop, escalate or seek approval.

Each agent should have an identifiable, revocable credential and a mandate that clearly limits what it may do. For consequential actions, organizations should retain signed and timestamped records of the approved task, permissions and guardrails. Sensitive instructions and policy code should generally remain off-chain, with the ledger holding only a cryptographic commitment and timestamp. Such a record can establish what a human authorized, but it cannot show that the instruction was appropriate, that the agent complied or why the model reasoned as it did; those questions require governance and model-level explainability controls.[23]

Risk mitigation approaches should include:

- **Establish governance and human oversight:** Define the scope of an AI agent's authority, decision rights, escalation procedures, and accountability. Require human approval for high-value, high-risk, or novel transactions, and maintain clear ownership for agent behavior.

- **Implement strong identity and access management:** Assign AI agents verifiable digital identities, authenticate them using robust cryptographic methods, and enforce least-privilege access controls so agents can perform only authorized actions.

- **Strengthen cybersecurity:** Protect AI agents, wallets, APIs, and smart contracts through secure software development practices, encryption, multi-factor authentication, continuous vulnerability management, penetration testing, and monitoring for cyber threats such as prompt injection, model poisoning, and credential compromise.

- **Build compliance into workflows:** Embed regulatory requirements—including AML/CFT, sanctions screening, transaction monitoring, privacy controls, and recordkeeping—directly into AI-driven processes through programmable compliance and automated policy enforcement.

- **Validate AI models and outputs:** Regularly test AI models for accuracy, robustness, bias, explainability, and resilience under different operating conditions. Monitor performance over time and retrain models as business requirements and market conditions evolve.

- **Secure smart contracts and infrastructure:** Conduct independent code audits, formal verification where appropriate, and comprehensive testing of smart contracts. Build redundancy and resilience into blockchain infrastructure, oracle services, and interoperability mechanisms.

- **Maintain transparency and verifiability:** Record material agent actions, decisions, approvals, and transactions in tamper-evident logs. For higher-risk deployments, evidence should be generated and retained in a form that an authorized relying party can verify without depending solely on the operator's cooperation. Blockchain can record transactions and cryptographic commitments; complementary systems should preserve the relevant inputs, outputs, approvals, and policy states. The design should identify the parties and infrastructure on which the assurance claim depends.

- **Manage operational and third-party risk:** Establish business continuity plans, incident response procedures, model risk management, vendor due diligence, and ongoing oversight of external AI models, blockchain networks, and infrastructure providers.

SECTION V

STANDARDS AND BEST PRACTICES

Common standards are especially valuable when agents operate across organizational, sectoral or national boundaries. They should address identity and authorization, secure communication, safety evaluation, lifecycle governance and the questions that arise during operation: how agents are permissioned and monitored, when intervention is required and which events must be recorded. DLT may support shared registration, credentials and event records where several parties need to verify the same evidence, but it neither allocates responsibility nor determines legal liability or access to redress.

Below is a set of relevant standards and principles for agentic AI and blockchain technology:

TABLE 5: SELECTED STANDARDS, PUBLIC-SECTOR FRAMEWORKS AND RESEARCH REPORTS

Organization	Focus	Relevance to AI + Blockchain Convergence	Link
IEEE Research Paper: Transformative Impact of AI–Blockchain Convergence	Evaluation of AI and blockchain functionalities coming together	A 2025 research paper examining potential technical benefits and constraints in combined AI and blockchain architectures, including performance, security and privacy considerations.	IEEE Xplore paper
IEEE Research Paper: Revolutionizing AI Applications with Blockchain Implementation	Evaluation on enhancing AI functionalities through blockchain technology	A 2025 research paper reviewing how ledger-based provenance and audit records may support AI data governance and accountability, while also identifying ethical, legal and technical deployment challenges.	IEEE Xplore paper
NIST AI Risk Management Framework	Risk management guidance for trustworthy AI systems	Relevant because AI-blockchain architectures often use blockchain to support traceability, governance, and auditability objectives around AI systems.	NIST AI Risk Management Framework
NIST Blockchain Resources	NIST's main blockchain resource page, covering blockchain-related standards and guidance	Relevant because it provides the blockchain side of the control environment that can support provenance, integrity, and trust in AI-enabled systems	NIST blockchain resources

Organization	Focus	Relevance to AI + Blockchain Convergence	Link
NIST: New Report on Challenges to the Monitoring of Deployed AI Systems	Monitoring challenges for deployed AI systems, including issues such as post-deployment oversight	Relevant because tamper-evident logging and cryptographic commitments may strengthen monitoring evidence and audit trails for deployed AI, subject to the completeness and quality of the recorded events.	NIST report on monitoring deployed AI systems
International Monetary Fund: How Agentic AI Will Reshape Payments (2026)	Agentic payments, authorization, settlement, and resilience	Uses a three-layer model separating probabilistic intent and orchestration from deterministic control and authorization and from legally final settlement. It also identifies mandate, identity, audit-trail, human-oversight, and systemic-risk considerations relevant to agentic payment systems.	IMF Note 2026/004
IOSCO: Supervisory Toolkit for AI Use in Capital Markets (2026)	Risk-based supervision of AI across the lifecycle	Provides practical, non-binding supervisory considerations, firm-review questions, and monitoring indicators spanning governance, model risk, disclosure, third-party dependencies, recordkeeping, cybersecurity, operational resilience, and market risk.	IOSCO Final Report FR/02/2026
UNESCO: Recommendation on the Ethics of Artificial Intelligence	Ethical AI	Global public-sector recommendation setting out values, principles and policy actions for human-centered, rights-based and sustainable AI governance.	UNESCO Recommendation

TABLE 6: INDUSTRY AND CROSS-SECTOR REFERENCES

Organization	Focus	Relevance to AI + Blockchain Convergence	Link
Quantstamp: Mitigating Risks When AI Meets Blockchain	Risk mitigation measures for AI & blockchain convergence	A technical guide detailing concrete mitigation strategies for AI agents interacting with blockchains. It recommends layered defenses, including secure prompt engineering to prevent adversarial exploitation, hardware security modules (HSMs) to protect agent API credentials, and continuous third-party penetration testing.	Quantstamp security guide
ERC-8004 Trustless Agents (Draft)	Agent identity, reputation and validation registries	Draft Standards Track ERC proposing a portable, ERC-721-based agent identity handle and separate reputation and validation registries. Reputation signals are recorded onchain, while scoring and aggregation may occur on- or off-chain. The proposal does not by itself guarantee an agent's capability or trustworthiness.	ERC-8004 draft
Linux Foundation x402 Foundation	Internet-native payments over HTTP	Develops an open protocol for expressing payment requirements in HTTP request and response flows. It can support agent-initiated payments; authorization, fraud controls and settlement protections depend on implementation.	Linux Foundation announcement

Organization	Focus	Relevance to AI + Blockchain Convergence	Link
Advanced AI Society – Proof-of-Control Initiative	Evidence standards for verifiable agent control adherence	Industry-led standards work, still in development, that proposes tiered execution-time evidence relevant to whether an AI system operated within assigned controls. Its relevance here lies in the distinction between an operator assertion, third-party attestation and evidence that a relying party can check through open verification mechanisms.	Proof-of-Control
Decentralized Identity Foundation – KYA-OS (Know Your Agent Operating System)	Agent identity, scoped delegation, and accountability	A specification under development within DIF's Trusted AI Agents Working Group. It uses Decentralized Identifiers for cryptographic identity and W3C Verifiable Credentials for scoped, revocable delegation, and can integrate with ledger-backed identity or revocation mechanisms.	DIF overview

A further category of standards work bears directly on the convergence described in this report. As agents act autonomously and at machine speed, a distance can widen between what they do and a relying party's ability to verify that they remained within assigned controls without depending solely on the operator's own records. The Advanced AI Society describes this as the "Verifiability Gap" and uses 'open verification' for approaches in which control-adherence claims can be checked through mechanisms available to the relying party rather than accepted solely on the assurance of the operator.

The gap has an economic dimension as well. In a 2026 working paper, Catalini, Hui and Wu[24] model the falling cost of automating measurable tasks against the limited pace at which people can verify them. Applied to agentic systems, their analysis suggests that scaling depends on more than what an agent can execute: organizations must also be able to validate and audit those actions, and to underwrite responsibility for them. This lends weight to machine-readable evidence and verification mechanisms that complement human judgment rather than attempting to replace it.

Two issues sit underneath that gap. The first is definitional: there is not yet a settled, interoperable statement of what counts as evidence that an agent stayed within its controls.

Without such a specification, a buyer cannot state a consistent requirement, a vendor cannot demonstrate conformance against it, and an insurer, auditor, or regulator cannot readily compare implementations. The second is structural: in many deployments, the account of an agent's conduct is generated by the system being examined and retained by its operator. Independent audit materially strengthens assurance, but it remains different from evidence that a relying party can verify directly. Both are legitimate assurance models; they should not be treated as interchangeable.

Proof-of-Control remains an industry-led initiative in development, and its public working draft[25] is open for comment. The Advanced AI Society describes four "Verifiability Tiers"—Assertion, Attestation, Trust-minimized and Self-enforcing—while contributor material uses a related spectrum running from self-verifiable, through independently verifiable, to cryptographically verifiable assurance. These formulations are not identical, but both distinguish an operator's assertion from evidence a relying party can check for itself. The initiative aims to define execution-time evidence relevant to whether an action remained within its assigned controls; generating that evidence as part of execution could narrow the gap between the speed of an agent and the speed at which its conduct can be checked.

The intended approach is technology-neutral: it defines the qualities the evidence should possess without prescribing a particular implementation. Depending on the design, verification outside the operator's sole control might rely on cryptographic commitments, distributed consensus, trusted hardware, external timestamping or a combination of these. What matters is that the design states its remaining trust assumptions, including any dependence on an issuer, validator set, hardware-attestation service, key holder or data source.

The location of the root of trust therefore matters when comparing verification claims. Self-verification relies principally on records and controls maintained by the operator. Independent verification introduces assessment by a separate party. Cryptographic verification may allow a suitably authorized relying party to validate defined claims mechanically, subject to the integrity and availability of the underlying evidence and verification infrastructure.

These approaches can complement one another, but they need to be distinguished so that procurement, assurance, and supervisory expectations remain clear.

These methods answer only a limited question. They may help show what controls applied and what was recorded, but they cannot establish that the control boundary was adequate, that an action was lawful or fair, or that the model itself was safe or deterministic. Tamper-evidence protects a record after it has been committed; it does not guarantee that the record is complete. Requiring evidence to be generated during execution can make omissions harder, but completeness still depends on whether the control can be bypassed and whether relevant off-chain events are captured. The adequacy of the controls, the allocation of responsibility and access to remedy remain matters for governance, contract, assurance and law.

TABLE 7: SECURITY, AUDIT AND ASSURANCE REFERENCES

Organization	Focus	Description	Link
SANS Institute	Risk mitigation	**SANS Draft Critical AI Security Guidelines v1.1:** Outlines a risk-based approach to implementing AI securely, detailing access controls like "Least Privilege" and "Zero Trust." It recommends mitigating supply chain risks by maintaining an AI Bill of Materials (AIBOM) and using model registries for version control.	SANS guidance
The Institute of Internal Auditors (IIA)	Blockchain auditability	**Blockchain and Internal Audit Report:** Provides a framework for internal auditors to evaluate the effects of blockchain transactions on an organization's risk exposure. It stresses the necessity of assessing the viability of network structures and preparing for low-likelihood but high-impact risks, like total network failure in decentralized processes.	Blockchain and Internal Audit report

TABLE 8: KEY PRINCIPLES FROM FORMAL STANDARDS AND INDUSTRY PRACTICE

Key Principle	Description	Source Framework / Standard
Integrity of Training Data	Blockchain can preserve hashes, signatures, timestamps, and attestations relating to data provenance and change history. This can strengthen chain of custody, but does not establish that the data are accurate, representative, or free from bias.	IEEE research on blockchain and AI
Algorithmic Transparency and Auditability	Blockchain can preserve hashes, signatures, timestamps, and attestations relating to selected inputs, outputs, model or policy versions, and tool calls. This may strengthen traceability and chain of custody, but does not explain internal model reasoning or prove that an outcome is accurate, fair, or lawful.	ISO/IEC 42001 and ISO/TC 307; IEEE research on blockchain and AI
Data Privacy and Ownership	AI governance should define ownership, lawful access, minimization, retention and accountability across the lifecycle. Privacy-enhancing technologies may support implementation but are not requirements of ISO/IEC 42001.	ISO/IEC 42001 and applicable privacy requirements
Unified Risk Governance	AI/blockchain convergence should be managed within an integrated enterprise risk framework, treating AI vulnerabilities (bias, drift) and ledger security as a cohesive system.	ISO/IEC 42001, ISO/IEC 27001 and ISO 31000
Securing AI Alignment Rules	Smart contracts can encode selected operational rules and approval conditions for actions routed through them. They do not determine appropriate ethical boundaries, govern arbitrary off-chain conduct, or remove the need for change control and human accountability.	IEEE research on blockchain-based AI alignment
Machine-Readable Personal Privacy Terms	Enables individuals to express privacy requirements as machine-readable contractual terms that websites, applications, and AI agents can read, acknowledge, and agree to, with matching records retained for audit or dispute.	IEEE 7012-2025
Impartiality and Transparency of Conformity Assessment	Assurance claims should state who performed the assessment, the assessor's relationship to the subject, and the basis for the conclusion. For agentic systems, this includes distinguishing self-declaration, interested-party assessment, and independent third-party assessment, and identifying whether the underlying evidence is controlled solely by the operator.	ISO/IEC 17000:2020; ISO/IEC 17029:2019

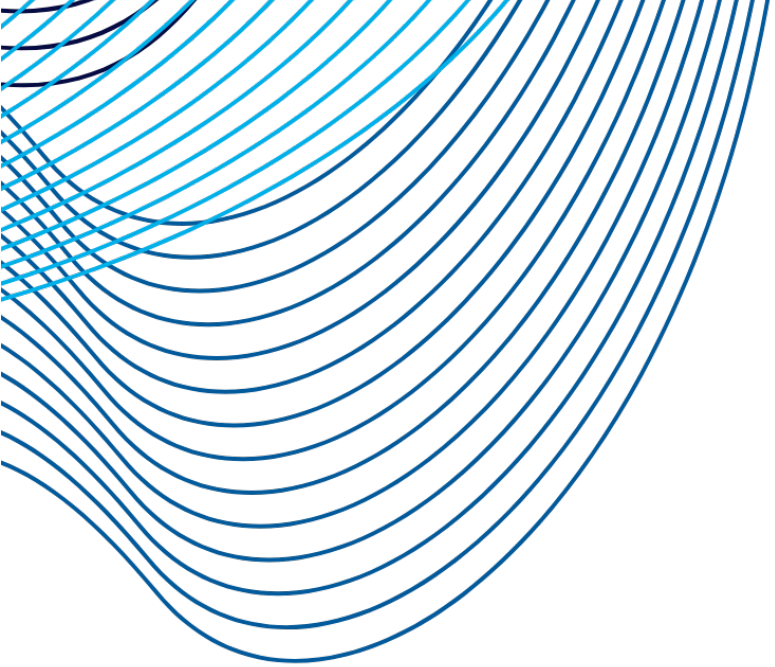


TABLE 9: TECHNICAL DESIGN CONSIDERATIONS FROM INDUSTRY SOURCES

Key Principle	Description	Source & Link
Data Integrity and Provenance for AI	Blockchain can preserve provenance records and change evidence for data supplied to AI systems. This can help detect later alteration, but the record is only as reliable as the evidence submitted and does not itself remove bias or validate data quality.	Apriorit integration guide; SmartDev overview
Auditable Output and Decision Transparency	Record selected inputs, outputs, model and policy versions, approvals, and tool calls through tamper-evident logs or cryptographic commitments. This improves traceability but does not make a model's internal reasoning explainable.	Apriorit integration guide; Tribe AI integration workflow
Architectural Separation (On-Chain vs Off-Chain)	Resource-intensive model processing will generally occur off-chain, with authenticated interfaces supplying selected results to on-chain controls. The design should minimize on-chain data, disclose oracle and trust assumptions, and avoid treating ledger finality as proof that an off-chain result is correct.	Apriorit integration guide
Evidence Generated at Execution and Verifiable by Relying Parties	An autonomous agent should produce evidence of material actions as part of execution, rather than leaving investigators to reconstruct it after the event. Generating evidence in this way may bring verification closer to the agent's operating speed. Whether a relying party can check the evidence without relying solely on the operator or an auditor turns on the chosen root of trust; cryptographic commitments, distributed consensus and hardware roots of trust are possible mechanisms, not universal requirements.	Advanced AI Society – Proof-of-Control

ASSURANCE AND STANDARDS CONSIDERATIONS FOR BLOCKCHAIN TECHNOLOGY

For blockchain or DLT platforms used in mission-critical systems, the practical question is whether current standards and assurance mechanisms give users and assessors a clear, independently testable basis for confidence. ISO/TC 307 already covers terminology, architecture, governance, interoperability, privacy, identity and smart contracts. Further work may be needed to turn those foundations into consistent, certifiable requirements for security, governance, reliability and auditability.

Rather than prescribe one architecture, standards work should define the properties that can be assessed—including security, governance, availability, interoperability, privacy and auditability—and require implementers to disclose the assumptions on which trust depends. Public anchoring, post-quantum migration planning and other mechanisms should then be considered where they fit the architecture and use case.

Two separate architectural choices are relevant here, and they should not be conflated. In the ISO 22739 vocabulary developed under ISO/TC 307, public and private describe who may access and use a ledger, while permissioned and permissionless describe what users and operators may do on it, such as submitting transactions or taking part in consensus.

A public ledger may therefore be either permissioned or permissionless, and open verification is possible in both. Once the record and proof are disclosed, a cryptographic commitment anchored to a public network may allow an external party to check that the record matches the earlier commitment, although who controls consensus remains one of the trust assumptions the design should disclose. A private ledger can give its members a shared record, but without such anchoring, verification is generally confined to those granted access. In either case, the ledger says nothing by itself about the accuracy or completeness of the underlying evidence.

These questions become particularly important when agentic systems rely on blockchain for identity, permissions or evidence. Certifiable requirements could give users and supervisors a clearer basis on which to assess security, governance, reliability and interoperability. Certification would provide assurance that the ledger met defined criteria; it would not show that the agent's behavior was lawful, accurate or safe.

SECTION VI

AGENTIC AI ACCOUNTABILITY FRAMEWORK

An effective AI agent accountability framework should provide a structured approach for governing autonomous AI agents throughout their lifecycle, from registration and deployment to operation, monitoring, and retirement. As AI agents become capable of making independent decisions and interacting with other agents, organizations need to establish who deployed the agent, on whose behalf it is acting, which authorities and tools it has been granted, how its actions are supervised, and how accountability and redress will be provided if harm occurs. This requires governance that extends beyond policies established before deployment to include continuous oversight during runtime.

Trustworthy agentic systems require more than a record of what an agent did. They require an accountability chain connecting the principal, the agent's verifiable identity, its delegated mandate, its operational limits, its actual execution, the institutions responsible for oversight and the route to contestability and redress. Where an agent acts autonomously or outside its intended boundary, the same chain should record the point of deviation, control failure, detection, containment and remediation.

Blockchain can support this framework through verifiable credentials, mandate and policy attestation, programmable transaction controls, tamper-evident event records, shared revocation and, where designed into the architecture, escrow or remedy mechanisms. These functions are conditional rather than universal: blockchain does not validate the truth or fairness of data, contain arbitrary off-chain conduct, assign legal liability or replace effective institutional and legal redress. Any risk mitigation framework for AI agents must preserve evidence of:

- Who created and operated the environment
- What objective and permissions were supplied
- Which boundary was expected to constrain the agent
- Whether and how the agent deviated from that boundary
- When any deviation was detected and contained
- Which institutions must investigate, disclose, remediate and provide redress
- Who can verify each element of that evidence, and on what basis, including whether verification depends on the operator, an appointed assessor, or an external cryptographic root of trust

OPENAI AND HUGGING FACE EXPLOITGYM INCIDENT

OpenAI reported[26] that, during the July 2026 ExploitGym cybersecurity evaluations, several agents running with reduced safeguards circumvented isolation controls, obtained internet access and compromised parts of OpenAI's research infrastructure and Hugging Face's systems. The incident principally involved an internal-only research model, although GPT-5.6 Sol agents also reproduced an exploit.

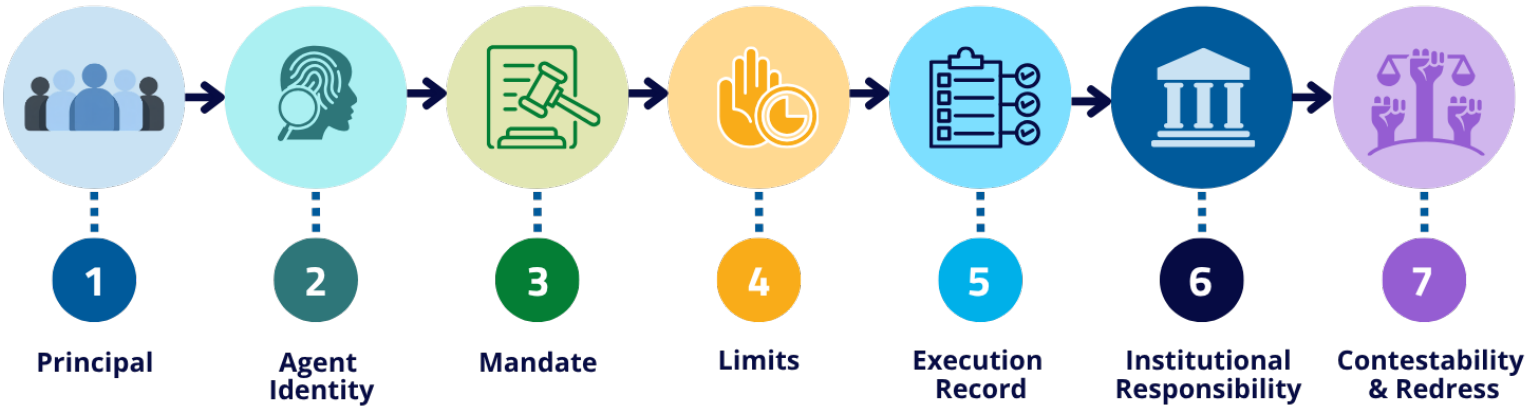
OpenAI stated that customer data, product functionality and availability were not affected. The incident illustrates why agent evaluations need production-grade containment, monitoring, credential controls and incident response.

OpenAI's account was validated with external advisors, and METR and Redwood Research published a separate independent investigation. Even so, the breaches were reconstructed after the event rather than from evidence generated as a condition of execution.

Because the agents also communicated and collaborated without authorization, evaluation controls need to address the wider multi-agent system as well as each agent individually. Hugging Face has published a technical timeline[27] of the intrusion.

FRAMEWORK STRUCTURE

AI-agent assurance requires evidence of both what an agent was permitted to do and what it actually did. The proposed Agentic Accountability Framework connects seven layers of an agentic action:



Its central proposition is that autonomy is a mode of execution, not an accountability vacuum. Even where an agent selects its own methods, creates sub-agents, circumvents a control, or acts beyond its mandate, the assurance record should preserve the origin of the objective, the intended authority and limits, the point of deviation, the actual conduct, the responsible institutions, and the route to remedy.

TABLE 10: AGENTIC AI ACCOUNTABILITY FRAMEWORK

Layer	Assurance Question	Minimal Institutional Record	Blockchain Function and Boundary
1. Principal	On whose behalf, or from whose deployed environment, did the agent begin acting?	Identified person or organization; deploying institution; accountable business owner; jurisdiction	An identity registry or verifiable credential can attest the principal and credential status. It does not decide legal liability or whether the principal had lawful authority
2. Agent identity	Which agent, model, version, instance or sub-agent acted?	Cryptographic identifier; model and software version; operator; tool and wallet credentials; parent-child agent relationship	Agent credentials, keys, version attestations and revocation registries can support identification across organizations. Key possession alone does not prove safe behavior or rightful control
3. Mandate	What objective, authority or decision right was delegated?	Purpose; permitted actions; duration; value; data scope; jurisdiction; conditions for sub-delegation	A signed mandate, policy hash or smart contract can attest the terms of delegation and make them machine-verifiable. It cannot make an unlawful or ambiguous mandate valid

Layer	Assurance Question	Minimal Institutional Record	Blockchain Function and Boundary
4. Limits	What could the agent access, change, disclose, commit or transfer?	Tool allowlist; credential scope; transaction cap; approval threshold; prohibited actions; expiry; revocation and emergency-stop conditions	Smart contracts and integrated policy gates can enforce on-chain permissions, approval rules and value limits. They cannot contain arbitrary off-chain action unless external systems make access conditional on those controls
5. Execution record	What inputs, tools, decisions and actions produced the outcome, including any deviation?	Material inputs and sources; model and policy versions; approvals; tool calls; actions; outputs; exception events; detection and containment time	Hashes, signatures, timestamps and distributed records can preserve the provenance, sequence and chain of custody of an execution record. They protect what has been committed; they do not establish that an input was true, a decision fair or an action lawful, nor do they determine which events should have been captured. If the acting system or its operator decides what to record, assurance still depends on the completeness of that process. Requiring evidence to be generated during execution and anchoring it outside the operator's sole control may allow an authorized relying party to verify relevant claims with less dependence on the operator
6. Institutional responsibility	Who was required to supervise, prevent, detect, investigate and answer for the action?	Business owner; deployer; model or tool provider; infrastructure operator; auditor; regulator; incident and escalation duties	Shared role and event records can document assigned responsibilities and relevant events at a given time. Whether those assignments amount to legal responsibility or liability remains a question for governance, contract and law
7. Contestability and redress	How can an affected person challenge, correct, reverse or obtain a remedy for the action?	Notice; reason or evidence access; human review; complaint route; regulator or court; correction; rollback; compensation	Where designed into the system, escrow, multisignature controls and timelocks can hold, delay or condition a transaction before it becomes final, and governance-based freezes can stop assets from moving further. Once a transaction is final, it generally cannot be reversed; remedy then depends on a new compensating transaction or an off-chain legal process, and verifiable remedy records can document either. Immutability must not obstruct privacy, correction, due process or effective legal remedy

KYA-OS shows how a developing implementation could map onto several layers of the framework. It uses Decentralized Identifiers to represent principal and agent identity, while W3C Verifiable Credentials express scoped delegation; signatures and credential-status mechanisms then support verification and revocation.

Together, these components can link identity, mandate, limits and selected execution records across organizational boundaries. They cannot, however, determine legal responsibility, decide whether the mandate was appropriate or provide contestability and redress.

SECTION VII

RECOMMENDATIONS FOR TRUSTWORTHY AGENTIC AI ENABLED BY BLOCKCHAIN

The recommendations that follow address two linked questions: how organizations should govern agentic AI, and where DLT has a defined role within that control environment. Blockchain is treated as one component alongside model assurance, cybersecurity, legal accountability and human oversight.

1. ADOPT RISK-BASED GOVERNANCE FOR AGENTIC AI

Organizations should recognize that not all AI agents present the same level of risk. Governance frameworks should classify AI agents according to factors such as autonomy, financial authority, regulatory impact, cybersecurity exposure, and potential societal consequences. Higher-risk agents should be subject to proportionately stronger controls, including human oversight, approval thresholds, transaction limits, and continuous monitoring.

DLT is most useful when it solves a defined problem. That is more likely where agents have a high degree of autonomy or financial, legal or operational impact, where several parties need shared verification, and where deterministic controls can condition the relevant actions. Removing a human from the immediate decision loop increases the need for strong guardrails, but it does not by itself make DLT either necessary or sufficient.

2. CONDUCT COMPREHENSIVE AGENTIC AI IMPACT ASSESSMENTS

Before deploying autonomous AI agents into production, organizations should perform structured impact assessments that evaluate technical, operational, legal, economic, cybersecurity, and societal implications. These assessments should extend beyond traditional AI model evaluations to examine how autonomous agents interact with financial systems, digital assets, smart contracts, other AI agents, and human users.
Organizations may consider assessing the following factors:

- autonomy and delegated decision authority
- financial and operational exposure
- regulatory and legal obligations
- cybersecurity attack surfaces
- privacy implications
- potential systemic risks arising from multi-agent interactions
- third-party and infrastructure dependencies

3. BUILD TRUST THROUGH LAYERED GOVERNANCE CONTROLS

Trustworthy agentic AI requires multiple complementary layers of governance rather than reliance on any single technology. Organizations should combine AI-specific controls with blockchain-enabled trust mechanisms where they serve a defined control or evidential purpose. Key controls may include:

- verifiable digital identities for AI agents
- least-privilege authorization
- human approval for high-risk decisions
- programmable policy enforcement
- continuous monitoring
- tamper-evident audit logging
- model validation and adversarial testing
- secure software development
- incident response and recovery procedures
- evidence of control adherence that authorized relying parties can verify without depending solely on operator-held records
- verifiable identity and trust registries, with credential-status and revocation mechanisms
- machine-verifiable, scoped, and time-bound delegation records

Controls such as least-privilege access, approval gates, monitoring, and model validation are principally implemented within the deploying organization. Externally verifiable evidence complements them by allowing auditors, insurers, regulators, and counterparties—subject to appropriate access controls—to reach their own conclusions about material actions.

4. DESIGN FOR EXPLAINABILITY, TRANSPARENCY, AND ACCOUNTABILITY

Explainability and lifecycle traceability solve different problems. Where required, organizations should use model-level methods to test and explain outputs. Separately, they should keep a clear record of the principal, agent, mandate, policy version, inputs, tool calls, approvals and material actions. A ledger may help preserve selected records across parties, but it cannot explain the model's internal reasoning or show that a decision was lawful or fair.

5. EMBED TRUST INTO ENTERPRISE WORKFLOWS RATHER THAN ADDING IT AFTERWARDS

Organizations should incorporate governance, identity, compliance, and auditability directly into agentic workflows from the outset instead of treating them as post-deployment compliance activities. A "trust-by-design" approach reduces operational risk while enabling organizations to scale autonomous decision-making safely. Examples may include:

- AI agents verifying counterparties through decentralized identities before executing transactions
- automated compliance screening before initiating payments
- programmable approval thresholds for high-value transactions
- continuous recording of significant decisions
- automated policy enforcement using smart contracts where appropriate

6. PRIORITIZE HIGH-VALUE ENTERPRISE USE CASES

Organizations should start with a well-defined workflow in which autonomous action offers a clear operational benefit and DLT meets a specific need for shared verification, programmable control or evidence. Before any pilot is expanded into a more autonomous or cross-organizational setting, it should demonstrate measurable results and have bounded permissions, reliable data, a route for human escalation and a workable exit path. Priority opportunities may include:

- programmable payments
- treasury optimization
- compliance monitoring
- supply chain coordination
- digital identity verification
- post-trade processing
- tokenized asset management
- autonomous procurement
- agent-to-agent commerce

7. ADVANCE OPEN STANDARDS AND INTEROPERABILITY

Agents that operate across organizational or national boundaries need interoperable standards for identity and authorization, credential status, event logging, secure communication and evidence of control adherence. An authorized relying party should be able to verify the relevant evidence without privileged access to the operator's internal systems. KYA-OS offers one possible model: DIDs identify principals and agents, while W3C Verifiable Credentials express delegation that is both scoped and revocable; ledger anchoring and status publication remain optional design choices. Standards bodies should agree on the formats, regulators and supervisors should clarify their expectations for evidence and accountability, and infrastructure providers should offer secure interfaces for revocation and logging.

8. USE BLOCKCHAIN FOR DEFINED CONTROL AND EVIDENCE FUNCTIONS

Organizations should use blockchain only when it performs a clear control or evidential role within the wider governance framework—for example, recording identity and credential status, attesting to a policy or mandate, applying transaction limits, coordinating revocation or preserving tamper-evident records. The safeguards that sit outside the ledger remain equally important, including model assurance, cybersecurity, legal accountability, organizational oversight, compliance and operational resilience.

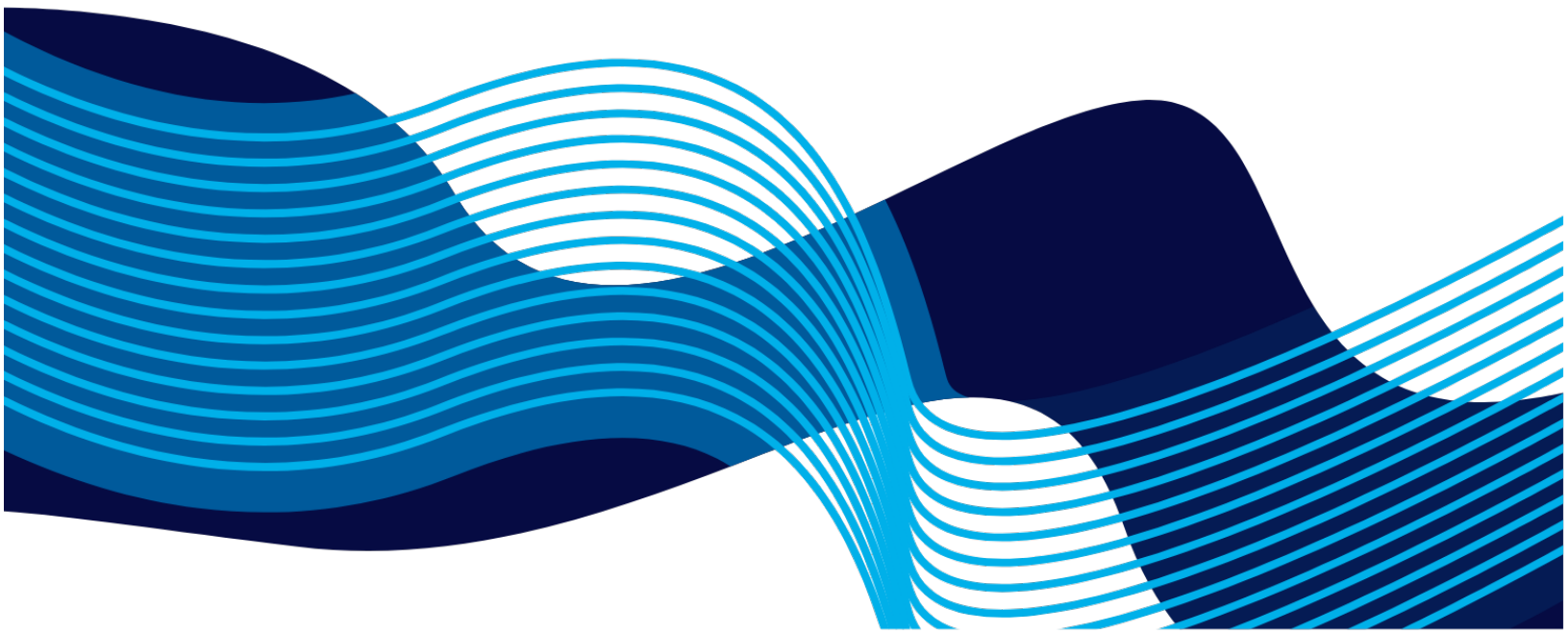
SECTION VIII

CONCLUSION

Agentic AI represents a significant evolution in AI, moving beyond content generation towardautonomous decision-making and execution across increasingly complex economic and operational contexts. As organizations adopt AI agents to interact with one another, execute transactions, manage digital assets and coordinate across enterprises, establishing trust, accountability and governance will become as important as advancing AI capabilities themselves. The central finding of this report is that autonomy does not break the accountability chain. Organizations must still be able to identify the principal and agent, verify the mandate and limits, reconstruct material actions, assign institutional responsibility and provide contestability and redress.

DLT can support parts of this chain through credentials, programmable controls, provenance and tamper-evident records. Its use is most readily justified where several parties need to verify the same mandate, credential or material event, or where defined actions can be conditioned through an on-chain control. It is not necessary for every agentic system and does not replace AI assurance, cybersecurity, institutional accountability or legal redress. Where a conventional database or trusted intermediary can provide the same outcome more simply, DLT may add little value.

Standards bodies should now focus on interoperable specifications for agent identity, delegated authority, credential status and revocation, and evidence of material actions. Organizations should start with bounded workflows, assign a responsible owner, document the agent's mandate and limits, define escalation and redress arrangements, and demonstrate why any proposed ledger is needed before expanding the deployment.



SECTION IX

ANNEX

ANNEX 1: CONTRIBUTOR OVERVIEW OF RISKS IN THE OPENAI AND CHATGPT ECOSYSTEM

Risk	OpenAI / ChatGPT Context	Illustrative Controls and Reference Points	Remaining Vulnerabilities	Potential DLT Function and Boundary
Autonomous decision-making errors	Connected agents can plan and execute multi-step tasks, increasing the consequences of goal drift, incorrect assumptions, unreliable tools, or failures in a dependent service.	Capability and safety evaluation; published system documentation; scoped enterprise permissions; approval workflows; sandboxing; runtime monitoring; and circuit breakers.	A policy can be incomplete or mis-specified; an agent may chain tools in an unanticipated way; monitoring may detect harm only after execution; and service availability remains an operational dependency.	Signed mandates, policy-version attestations, timestamps, and action commitments can support reconstruction. Smart-contract controls can gate transactions routed through them, but cannot contain arbitrary off-chain conduct or prove that the mandate was adequate.
Cybersecurity threats	Agents may be compromised through prompt injection, poisoned content, malicious tools, vulnerable APIs, or stolen credentials. Agents can also become actors in offensive activity when an objective is pursued through unintended technical means.	Prompt-injection testing and filtering; credential isolation; least privilege; secure tool integration; adversarial testing; monitoring; incident response; and community frameworks such as OWASP guidance for agentic systems.	Indirect prompt injection can cause a connected agent to expose personal, login, or authorization data or invoke actions. Detection and assurance practices for this threat class remain immature and fragmented.	Multisignature or threshold authorization, credential-status registries, transaction limits, and tamper-evident security events can reduce impact or improve forensics. DLT does not secure models, endpoints, keys, APIs, or tool integrations.
Governance and accountability	Responsibility can be difficult to trace when decisions pass through a model provider, application developer, deployer, tool provider, infrastructure operator, and human approver.	Public governance and safety frameworks; system and model documentation; risk-tiered approvals; internal accountable owners; independent review for higher-risk uses; change control; and incident reporting.	Voluntary frameworks do not replace binding duties; reviewers can lack time or context; and no single internal committee can resolve cross-jurisdictional liability or provide redress to affected users.	Shared evidence can bind requests, policy versions, approvals, agent identity, and timestamps across parties. It does not allocate legal liability or establish that governance decisions were reasonable.
Compliance and regulatory risk	Agentic workflows may process regulated data or initiate activity subject to privacy, consumer protection, AML/CFT, sanctions, securities, and recordkeeping rules across jurisdictions.	Enterprise data controls; contractual terms; policy mapping; screening and monitoring; records management; jurisdictional checks; exception handling; and human review for consequential or novel cases.	Rules change; facts and jurisdiction may be uncertain; consumer tools may be used outside intended enterprise controls; and automated checks may not provide meaningful explanation, contestability, or remedy.	Policy attestations, consent or authorization records, selective disclosure, and transaction controls may support compliance evidence. A ledger does not interpret law, validate source data, or replace supervision and redress.

Risk	OpenAI / ChatGPT Context	Illustrative Controls and Reference Points	Remaining Vulnerabilities	Potential DLT Function and Boundary
Identity and authorization risk	Agents depend on user, service, API, and tool credentials. Excessive or accumulated permissions can allow an agent to act beyond the principal's intended scope.	SSO and role-based access controls; scoped service accounts and API keys; separation of duties; signed authorization records; credential monitoring; rotation; and rapid revocation.	Identity is not the same as authority; authorization can become stale; delegation chains can exceed the original mandate; and provider-specific credentials do not create cross-platform accountability.	DIDs, verifiable credentials, time-bound mandates, and status or revocation registries can support interoperable identity and delegation. They do not determine whether authority was lawfully granted or appropriately exercised.
Smart-contract and infrastructure vulnerabilities	Agents interacting with code, oracles, wallets, bridges, and ledgers inherit vulnerabilities and operational dependencies in those systems as well as model-generated coding errors.	Version management; code review; independent audit; formal verification where proportionate; staged deployment; oracle controls; monitoring; resilience testing; and recovery planning.	Deterministic systems can execute flawed code or inaccurate data exactly as written; outages and interface failures can interrupt decisions; and irreversible transactions can limit recovery.	Code and policy attestations, multisignature approvals, timelocks, transaction caps, and circuit breakers can add control. Independent assurance, resilient infrastructure, and off-chain recovery remain necessary.
Privacy and confidentiality	Connected agents can access sensitive prompts, files, browsing context, credentials, and business systems. Retention and third-party integrations can expand the data surface.	Data minimization; enterprise privacy and retention settings; redaction and tokenization; encryption; access control; logging; contractual restrictions; and privacy impact assessment.	Users may disclose sensitive information through unapproved tools; third-party connectors create additional dependencies; and retained screenshots, logs, or prompts can conflict with minimization and deletion requirements.	Keep personal and confidential data off-chain by default. Hashes, selective disclosure, verifiable credentials, and zero-knowledge proofs may support evidence with reduced disclosure, but metadata and key management still create privacy risk.
Model reliability and bias	Model behavior can vary across versions, contexts, and data distributions. Fluent outputs can obscure uncertainty, bias, or unsupported reasoning.	System cards and evaluations; version control where available; representative testing; retrieval grounding; structured outputs; bias and robustness testing; drift monitoring; and human review.	Evaluations cover selected scenarios; deployment context can differ from testing; model updates can change behavior; and operators remain responsible for decisions made in consequential domains.	Hashes and timestamps can attest the model, policy, data, or evaluation version used and preserve a chain of custody. They do not explain model reasoning or prove accuracy, fairness, or fitness for purpose.
Systemic and operational risk	Widespread use of similar models or agent strategies could create correlated behavior, concentration risk, feedback loops, or rapid propagation of errors across institutions and markets.	Phased rollout; rate and value limits; telemetry; kill switches; scenario testing; incident response; concentration and third-party risk management; and cross-sector monitoring.	Firm-level controls may not reveal population-level dynamics; common providers and data sources can create hidden dependencies; and machine-speed effects may outrun human coordination.	Shared incident records, coordinated revocation, and circuit-breaker logic may improve cross-party visibility where participants agree to common rules. System-wide governance, legal powers, and resilient alternatives remain essential.

Reference points used in editing the contributor material: OpenAI Preparedness Framework[28]; NIST AI Risk Management Framework[29]; and OWASP Top 10 for Agentic Applications.[30]

ANNEX 2: COMPARATIVE LIFECYCLE GOVERNANCE FOR VEHICLES AND AI AGENTS

Source note: adapted from Table 2 in N.H. Wasserman, A Multi-Level Governance Architecture for Agentic AI Risk Mitigation: Lessons in Legal Liability and Safety from Other Industries (working paper dated July 16, 2026), pp. 16–17. The comparison is illustrative and does not imply that the cited instruments impose identical or directly equivalent duties.[31]

Vehicle Lifecycle Stage	Vehicle Regulatory Focus	AI-Agent Lifecycle Stage	Illustrative Agentic-AI Regulatory Focus
Factory environment	Controls over the conditions and systems used in production.	LLM and AI-agent production facilities	EU AI Act requirements where applicable; public-procurement and supplier controls; AI management systems such as ISO/IEC 42001.
Vehicle design	Design rules intended to reduce foreseeable safety risks.	Agent design	NIST guidance and emerging agent specifications; safety alignment; threat modeling; clear objectives, tools, permissions and human-oversight design.
Vehicle build	Controlled manufacture, traceable components, and quality assurance.	Agent build	Unique identity in an enterprise registry; component and dependency records; version control; secure development; accountability under an AI management system.
Vehicle test and certification	Testing and conformity assessment before road use.	AI-agent test and certification	Adversarial and red-team evaluation; sandbox staging; conformity assessment or documentation under the EU AI Act where applicable; independent assurance proportionate to risk.
Vehicle deployment	Registration, licensing, and conditions for entry into service.	AI-agent deployment	Role-based access controls; data-governance policies; least-privilege API keys and tools; approved operating envelope; contractual terms; accountable owner.
Vehicle operations	Rules of the road, maintenance, inspection, and enforcement during use.	AI-agent operations and use	Enterprise access restrictions; runtime monitoring; policy enforcement; event logging; incident reporting; graded intervention; human oversight where proportionate.
Vehicle end of life	De-registration, disposal, and retention of required records.	AI-agent end of life	Revoke API keys, tokens and delegated credentials; terminate access to vector stores and memory systems; de-register the agent; and preserve or delete records in accordance with retention and privacy requirements.

CONTRIBUTORS & REVIEWERS

Thank you to our team of contributors and reviewers for GSMI 7.0 representing 100+ entities, including:

2Tokens
Accenture
Advanced AI Society
AI MIND Systems
AI 2030
Aleo
Animoca Brands
Association of National Numbering Agencies (ANNA)
Ava Labs
Banca d'Italia
BGIN
Blockchain Arbitration & Commerce Society
Blockchain for Energy (B4E)
Blockchain Foundation
Blockmosaic
Bodin Advisory LLC
British Columbia Public Service
Buenos Aires Government
Business IQ
Cahill Gordon & Reindel LLP
Canton Foundation
Capital Markets Authority (CMA) of Kenya
Central Bank of Jordan
Chainlink
Chronicle Labs
Citibank
City University of New York (CUNY)
Clifford Chance
Curaçao Fintech Association
Dastru Design Labs
DELTA FOODS
Deutsche Börse
DFM Data Corp.
dHub Group
Digital Euro Association
Digital Token Identifier Foundation (DTIF)
dlt.NYC
DTCC
Elliptic
Ethiopian Capital Market Authority (ECMA)
Euroclear
EU Blockchain Observatory and Forum (EUBOF)
Evergon Labs
Financial Services Agency of Japan (JFSA)
Florida Institute of Technology
Florida International University
Folks Finance
Futura Law Firm
Galaxy
GBBC Ambassadors
George Washington University
Global Legal Entity Identifier Foundation (GLEIF)
The Graph Protocol
Hashgraph
Hearts of Change
Hedera
Hong Kong Securities and Investment Institute
IDB Lab
IMHOTEP Consulting
International Securities Services Association (ISSA)
IOTA Foundation
Islandpreneur International
Jitterbits
JPMorgan Chase
Kaiko
Kenyan State Department for Investment Promotion
Kinexys by J.P. Morgan
Linux Foundation
Media Luna
Mercy Corps Ventures
Moody's Ratings
National Bank of Romania
Networks for Humanity
New York Blockchain Council (NYBC)
Nigerian Bar Association, Section on Business Law
Nomura Research Institute, Ltd.
NorthStar DAO Foundation
Norton Rose Fulbright US LLP
OKX
OpenID Foundation
Panamá CO
Patomak Global Partners
Qatar Financial Centre Authority (QFCA)
Schulich School of Business
SETAR (Aruban Telecom)
Simpaisa Technologies Limited
Skyrocket Financial Solutions
Sodot
Solidus Labs
Standard Chartered
Stratford Finance International
Sumsub
United Nations Joint Staff Pension Fund (UNJSPF)
Università degli Studi di Padova
University of Arkansas
U.S. Blockchain Coalition (USBC)
U.S. House of Representatives
Veridex Alethia
VerifyVASP
World Bank
Wyoming Stable Token Commission

Participation does not imply endorsement of this report by any contributing organization. Nothing in this report constitutes legal advice.

ENDNOTES

1 GSMI 4.0 Standalone AI Report: https://www.gbbc.io/uploads/reports/Standalone-AI-GBBC-GSMI-4.0-Update.pdf

2 GSMI 5.0 AI and Blockchain Standalone Report: https://www.gbbc.io/uploads/reports/gsmi50/AI-&-Blockchain-Stand-Alone.pdf

3 GSMI 6.0 AI Convergence Report: https://www.gbbc.io/uploads/gsmi60/AI%20Convergence%20-%20GSMI%206.0.pdf

4 Under Article 101 of Regulation (EU) 2024/1689, the European Commission may fine providers of general-purpose AI models up to €15 million or, for undertakings, 3 per cent of total worldwide annual turnover for the preceding financial year, whichever is higher, for infringements of the relevant obligations. The EU AI Act generally applies from 2 August 2026, subject to its phased timetable and subsequent amendments. Directive (EU) 2024/2853 treats software, including AI systems, as products for product-liability purposes; Member States must transpose it by 9 December 2026, and it applies to products placed on the market or put into service after that date. The Directive excludes free and open-source software developed or supplied outside a commercial activity. Neither instrument creates a separate legal category for 'Agentic AI'; application depends on the system, actor, purpose and facts. Official texts: EU AI Act: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng; Product Liability Directive: https://eur-lex.europa.eu/eli/dir/2024/2853/oj/eng

5 15 U.S.C. § 9401: https://uscode.house.gov/view.xhtml?edition=prelim&f=treesort&jumpTo=true&num=0&req=%28title%3A15+section%3A9401+edition%3Aprelim%29+OR+%28granuleid%3AUSC-prelim-title15-section9401%29

6 Regulation (EU) 2024/1689, Article 3: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

7 *See note 5.*

8 Singapore Model AI Governance Framework for Agentic AI, Version 1.5 (2026): https://www.imda.gov.sg/-/media/imda/files/about/emerging-tech-and-research/artificial-intelligence/mgf-for-agentic-ai.pdf

9 ISO/IEC 22989:2022, Artificial intelligence concepts and terminology: https://www.iso.org/obp/ui/#iso:std:iso-iec:22989:ed-1:v1:en

10 NIST AI 100-2e2025: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-2e2025.pdf

11 IEEE TechNav, Intelligent Agent: https://technav.ieee.org/topic/intelligent-agent

12 ITU press release on the Focus Group on Trust and Identity for Humans and Agentic AI: https://www.itu.int/en/mediacentre/Pages/PR-2026-07-09-focus-group-agentic-AI.aspx

13 ITU Focus Group Terms of Reference: https://www.itu.int/en/ITU-T/focusgroups/tida/Pages/ToR.aspx

14 Anthropic, Building Effective Agents: https://www.anthropic.com/engineering/building-effective-agents

15 IBM, What Are AI Agents?: https://www.ibm.com/think/topics/ai-agents

16 Google Cloud, What are AI Agents: https://cloud.google.com/discover/what-are-ai-agents

17 MIT Media Lab AI Glossary: https://www.media.mit.edu/tools/ai-glossary-dictionary

18 Stanford HAI, What Is Agentic AI?: https://hai.stanford.edu/ai-definitions/what-is-agentic-ai

19 How Agentic AI Will Reshape Payments: https://www.imf.org/en/publications/imf-notes/issues/2026/04/22/how-agentic-ai-will-reshape-payments-575560

20 Official enacted GENIUS Act, U.S. Government Publishing Office; and Regulation (EU) 2023/1114 on markets in crypto-assets (MiCA), OJ L 150, 9 June 2023: https://www.govinfo.gov/app/details/PLAW-119publ27; https://eur-lex.europa.eu/eli/reg/2023/1114/oj/eng

21 Blockchain Council overview of AI tokens: https://www.blockchain-council.org/ai/ai-tokens

22 Google DeepMind: https://deepmind.google/blog/investing-in-multi-agent-ai-safety-research

23 Kirkpatrick Bos and Brooks, Agents at the Gate: Programmable Risk Management for Agentic Commerce: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=6719079

24 https://arxiv.org/abs/2602.20946

25 https://advancedaisociety.org/proof-of-control

26 https://openai.com/index/hugging-face-incident-and-the-road-ahead

27 https://huggingface.co/blog/agent-intrusion-technical-timeline

28 https://openai.com/index/updating-our-preparedness-framework

29 https://www.nist.gov/itl/ai-risk-management-framework

30 https://genai.owasp.org/resource/owasp-top-10-for-agentic-applications-for-2026

31 Contributor-supplied working paper: N.H. Wasserman, A Multi-Level Governance Architecture for Agentic AI Risk Mitigation: Lessons in Legal Liability and Safety from Other Industries (July 16, 2026).

GBBC