An Agentic AI Framework for Distributed Quality Control in Industry 4.0

Ilenia Ficili, Giuseppe Tricomi, Luca D'Agati, Francesco Longo, Giovanni Merlino, Antonio Puliafito
Department of Engineering (DI), University of Messina, Messina, Italy
{ilficili, gtricomi, ldagati, flongo, gmerlino, apuliafito}@unime.it

***Abstract*—This paper proposes a decentralized and auditable framework for quality control in Industry 4.0 manufacturing environments. The architecture combines unsupervised anomaly detection, Distributed Ledger Technology (DLT), and Agentic AI to move beyond conventional pipelines that mainly generate isolated alerts. Detected anomalies are converted into structured events and notarized on a private permissioned ledger, providing a trusted record for autonomous decision-making. Agentic AI components monitor these events, assess their operational relevance through risk-aware reasoning, and generate mitigation decisions that can be executed by industrial control systems. By linking perception, ledger-based traceability, autonomous reasoning, and execution, the framework supports more responsive, transparent, and coordinated defect management across heterogeneous cyber-physical systems.**

***Index Terms*—Unsupervised anomaly detection, Distributed Ledger Technology, Agentic AI, Industry 4.0, Cyber-Physical Systems.**

I. Introduction

In modern Industry 4.0 manufacturing environments, operational responsiveness in quality control, zero-defect manufacturing, and predictive maintenance represents a critical operational requirement. Traditionally, Anomaly Detection (AD) pipelines operate as unidirectional, passive, and highly siloed systems. A typical machine learning model identifies a deviation from the standard baseline and issues a binary flag or a localized alert. This process inherently shifts the subsequent burden of diagnostic analysis, root-cause investigation, and mitigation strategies onto human operators or centralized, monolithic supervisory control systems.

While effective in legacy or legacy-isolated operations, this traditional setup introduces severe bottlenecks within next-generation smart manufacturing lines. These environments are increasingly characterized by massive automation, high-throughput production rates, and a compelling need for seamless orchestration across highly independent, heterogeneous subsystems (e.g., robotic arms, conveyor belts, and autonomous guided vehicles). Relying on human-in-the-loop interventions introduces unpredictable cognitive latency, error-prone decision-making under stress, and an absolute lack of deterministic logging. Furthermore, centralized industrial control systems (such as traditional MES or SCADA) act as single points of failure and lack the cryptographic mechanisms required to satisfy strict modern regulatory auditing, cross-enterprise accountability, and supply chain transparency mandates. A major unresolved challenge in these interconnected cyber-physical systems (CPS) is the establishment of decentralized trust. When multiple autonomous components from different vendors must cooperate and react to operational failures, a shared, unalterable, and reliable record of events is strictly necessary to prevent data falsification and guarantee auditability. To overcome these structural constraints, this work introduces a conceptual paradigm shift, formalizing the evolution of the industrial operational loop from a localized, human-dependent sequence toward an interconnected, distributed, and trust-aware ecosystem:

$$\text{detection} \longrightarrow \text{DLT notarization} \longrightarrow \text{autonomous decision} \longrightarrow \text{physical action} \quad (1)$$

The core objective of this architectural Proof of Concept (PoC) is the native convergence of edge-level perception, immutable distributed memory, multi-agent reasoning, and automated physical execution into a single, cohesive, trust-aware architecture. To summarize, the main contributions of this extended abstract are as follows: **i)** we propose a decentralized, closed-loop conceptual framework that transforms unsupervised anomaly detection outputs into actionable, traceable industrial decisions; **ii)** we integrate a private permissioned distributed ledger layer (Hyperledger Fabric) to act as a secure coordination substrate and immutable memory for autonomous systems; **iii)** We formalize a risk-based mathematical decision logic for autonomous Agentic AI components, bridging the gap between raw data perception and verifiable physical mitigation without human intervention.

II. Related Work

Visual inspection and automated defect localization have reached strong results due to the availability of high-quality industrial benchmarks and increasingly effective unsupervised anomaly detection methods. Datasets such as *MVTec AD 2* provide realistic testbeds for evaluating defect detection and localization under heterogeneous manufacturing conditions [1]. However, most existing work remains primarily focused on detection accuracy and localization metrics, giving limited attention to how anomaly outputs can be integrated into downstream decision-making, traceability, and control workflows.

In parallel, the convergence of Artificial Intelligence and Distributed Ledger Technology (DLT) has been investigated as a way to improve data integrity, accountability, and trust in IoT and industrial IoT scenarios [2]. Permissioned blockchain platforms such as Hyperledger Fabric are particularly relevant for industrial environments because they support controlled membership, modular consensus, and enterprise-oriented deployment models [3]. More recently, Agentic AI research has introduced software entities capable of combining perception, reasoning, planning, and action, as reflected in reasoning-and-

acting paradigms and recent surveys on autonomous LLM-based agents [4].

III. Proposed Architecture and Methodology

The proposed framework is structured across four functional layers designed to handle the transition from raw sensor telemetry to certified physical mitigation.

Perception Layer: The primary layer interfaces directly with the physical environment. Powered by an unsupervised anomaly detection model validated against the industrial scenarios of MVTec AD 2, it ingests video streams or multi-modal sensor telemetry. The model outputs a structured payload containing: a normalized anomaly score ($S_{ad} \in [0, 1]$), localization data (e.g., pixel-level segmentation maps or bounding boxes), and model metadata including versioning and inference confidence metrics.

DLT Layer The DLT layer serves as the backbone of architectural trust. This architecture is conceptualizedupon a private permissioned ledger utilizing Hyperledger Fabric; this is the best candidate due to its modular architecture and support for private data collections, which shield sensitive industrial property. The ledger, acting as an immutable event registry, ensure computational efficiency, heavy imagery data are retained off-chain; the ledger instead seals the cryptographic SHA-256 hashes of the files alongside operational metadata, guaranteeing that data cannot be altered retroactively.

Agentic AI Layer and Reasoning Logic This layer hosts autonomous, intelligent software agents that continuously monitor the Hyperledger Fabric state updates. Upon the notarization of a new anomaly event, the corresponding agent initiates an internal *Perceive-Reason-Act-Commit* execution cycle.

To formalize the agent's decision-making process without relying on arbitrary heuristics, the agent computes an operational risk index (R). This index mathematically weights the real-time anomaly score against the localized asset critical factor and recent historical defect density:

$$R = (S_{ad} \times W_m) + \alpha \cdot \left(\frac{N_f}{\Delta t} \right) \tag{2}$$

where $W_m \in [1, 5]$ represents the asset importance weight, N_f is the number of recorded failures within the moving time window Δt, and α is a scaling normalization coefficient. If R exceeds a predefined critical safety threshold (R_{crit}), the agent automatically formulates a mitigation strategy (e.g., emergency stops, component diversion) and commits the cryptographically signed decision back to the ledger.

Execution Layer The final layer operationalizes the validated logic. Edge controllers, PLCs, SCADA, or Manufacturing Execution Systems (MES) intercept the signed decision directly from the trusted ledger, executing the physical or digital commands deterministically, thereby closing the control loop.

IV. Data Interface and Architectural Design

Since this work serves as a conceptual Proof of Concept, the data structures that govern the boundaries between the independent layers must be strictly defined to ensure future interoperability. Fig. 1 relies on standard, lightweight serialization formats to minimize serialization overhead at the edge.

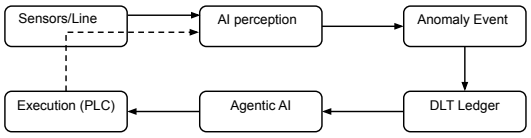

Fig. 1. End-to-end operational flow of the decentralized closed-loop system.

LISTING A

```
{"event_id": "evt_9942",
"machine_id": "rob_arm_04",
"anomaly_score": 0.92,
"threshold": 0.80,
"status": "detected",
"data_hash": "a4f2b9..."}
```

LISTING B

```
{"event_id": "evt_9942",
"agent_id": "ag_maint_01",
"decision": "trig_emerg_stop",
"reason": "Risk score R
exceeded critical threshold",
"timestamp": 1779224850}
```

Fig. 2. A: Conceptual anomaly event payload for ledger entry; B: Conceptual decision payload committed by the agent.

Listing A in Fig. 2 formalizes the interface data model emitted by the Perception Layer node. This structure ensures that the DLT layer registers the minimum amount of structured telemetry required for tracking, including the data hash for off-chain verification. Consequently, Listing B in Fig. 2 maps the explicit decision artifact committed back to Hyperledger Fabric by the Agentic AI node after evaluating the risk index R. By analyzing this conceptual data distribution, we can objectively evaluate the systemic bottlenecks inherent to the proposed design: **i)** Latency-Throughput Trade-off: Even within private permissioned DLT infrastructures like Hyperledger Fabric, block consensus mechanisms introduce a processing delay. This latency is incompatible with the sub-millisecond, hard real-time execution loops often required by fast-paced physical factory components; **ii)** State Congestion: High-frequency operational checks generate thousands of data transactions per second. Recording every minor deviation on the ledger ledger would trigger computational scalability issues, reinforcing the need for a strict edge-side threshold filter before DLT submission;

V. Conclusion and Future Work

This work demonstrates a solid architectural study that shifts industrial quality monitoring from a passive state into an inherently auditable, secure, and autonomous governance model. As this framework is currently established as a theoretical PoC, future work will focus directly on empirical implementation and physical validation.

Acknowledge

This work is partially supported by SquadraIoT company.

References

[1] L. Heckler-Kram, J.-H. Neudeck, U. Scheler, R. König, and C. Steger, "The MVTec AD 2 dataset: Advanced scenarios for unsupervised anomaly detection," *International Journal of Computer Vision*, vol. 134, no. 4, 2026, doi: 10.1007/s11263-026-02743-0.

[2] S. Latif, Z. Idrees, Z. e Huma, and J. Ahmad, "Blockchain technology for the industrial Internet of Things: A comprehensive survey on security challenges, architectures, applications, and future research directions," *Transactions on Emerging Telecommunications Technologies*, vol. 32, no. 11, 2021, Art. no. e4337, doi: 10.1002/ett.4337.

[3] E. Androulaki *et al.*, "Hyperledger Fabric: A distributed operating system for permissioned blockchains," in *Proc. 13th EuroSys Conf.*, 2018, pp. 1–15, doi: 10.1145/3190508.3190538.

[4] L. Wang *et al.*, "A survey on large language model based autonomous agents," *Frontiers of Computer Science*, vol. 18, no. 6, Art. no. 186345, 2024, doi: 10.1007/s11704-024-40231-1.