Article

Not peer-reviewed version

Human-Centric Zero Trust Identity Architecture for the Fifth Industrial Revolution: A JEPA-Driven Approach to Adaptive Identity Governance

Jovita T. Nsoh *

Posted Date: 20 March 2026

doi: 10.20944/preprints202603.1618.v1

Keywords: adaptive identity governance; critical infrastructure; fifth industrial revolution; human-centric security; identity and access management; industry 5.0; JEPA; operational technology; privacy-preserving authorization; zero trust architecture

 doi:10.20944/preprints202603.1618.v1

Article

Human-Centric Zero Trust Identity Architecture for the Fifth Industrial Revolution: A JEPA-Driven Approach to Adaptive Identity Governance

Jovita T. Nsoh

Department of Engineering Technology, Cullen College of Engineering, University of Houston, Houston, TX 77204, USA; jtnsoh@uh.edu

Abstract

The Fifth Industrial Revolution (Industry 5.0) foregrounds human–machine collaboration, sustainability, and resilience as organizing principles for next-generation cyber-physical systems. Yet the identity and access management (IAM) architectures inherited from Industry 4.0 remain perimeter-centric, policy-static, and blind to the behavioral dynamics of human–AI teaming. This paper introduces the Human-Centric Zero Trust Identity Architecture (HC-ZTIA), a novel framework that repositions identity as the adaptive control plane for Industry 5.0 environments. HC-ZTIA integrates three mutually reinforcing innovations: (1) a Joint Embedding Predictive Architecture (JEPA)-driven Behavioral Identity Assurance Engine (BIAE) that learns abstract world models of operator and machine-agent behavior to perform continuous, context-aware identity verification without relying on raw biometric surveillance; (2) a Privacy-Preserving Adaptive Authorization Protocol (PP-AAP) employing zero-knowledge proofs and federated policy evaluation to enforce least-privilege access across human, non-human, and hybrid identity classes while satisfying data-minimization mandates; and (3) a Resilience-Oriented Trust Degradation Model (RO-TDM) that guarantees fail-safe identity governance under adversarial, degraded, or disconnected operating conditions characteristic of operational technology (OT) and critical infrastructure. The framework is grounded in the Agile-Infused Design Science Research Methodology (A-DSRM) and formally extends NIST SP 800-207 and the CISA Zero Trust Maturity Model by addressing five identified gaps in human-centric identity governance. We present the formal system model, threat model, architectural specification, and a multi-scenario evaluation spanning energy-sector OT, smart manufacturing, and vehicle-to-everything (V2X) environments. Simulation results, validated through Monte Carlo trials with 95% confidence intervals, demonstrate that HC-ZTIA reduces identity-related breach exposure by 73.2% (±4.1%) while maintaining sub-200 ms authorization latency, offering a principled bridge between Zero Trust rigor and Industry 5.0 human-centricity.

Keywords: adaptive identity governance; critical infrastructure; fifth industrial revolution; human-centric security; identity and access management; industry 5.0; JEPA; operational technology; privacy-preserving authorization; zero trust architecture

1. Introduction

The transition from Industry 4.0 to Industry 5.0 marks a paradigmatic reorientation of industrial systems: from automation-maximizing efficiency to human-centric collaboration, sustainability, and societal resilience [1,2]. Where the Fourth Industrial Revolution sought to connect machines, the Fifth seeks to harmonize human creativity with machine intelligence, placing the well-being, agency, and cognitive augmentation of human operators at the center of system design [3]. This shift carries profound implications for cybersecurity. Legacy security architectures, designed around perimeter-based trust models and static role-based access control, are structurally inadequate for environments

 doi:10.20944/preprints202603.1618.v1

in which humans and AI agents dynamically co-create, share decision authority, and operate across fluid trust boundaries [4,5].

Identity and Access Management (IAM), long treated as an administrative function subordinate to network security, must now be elevated to the status of the *control plane* of Industry 5.0 security [6]. In Zero Trust Architecture (ZTA), as codified in NIST SP 800-207 [7], every access request is continuously verified irrespective of network location. Yet current ZTA implementations exhibit five critical gaps when confronted with Industry 5.0 requirements:

Gap 1 (Behavioral Blindness): Existing ZTA identity verification relies on static credentials, multi-factor authentication tokens, and point-in-time device posture checks. It lacks continuous behavioral modeling that captures the evolving cognitive and operational patterns of human–machine teams [8,9].

Gap 2 (Identity Class Fragmentation): Human identities, non-human identities (NHIs) such as service accounts and API keys, and emergent hybrid identities (AI agents acting with delegated human authority) are governed by separate, uncoordinated policy silos. Industry 5.0 demands unified identity governance across all classes [10,11].

Gap 3 (Privacy–Utility Antagonism): Human-centric environments generate sensitive behavioral, biometric, and cognitive telemetry. Current ZTA frameworks offer no mechanism to perform continuous verification while simultaneously enforcing data minimization and privacy-by-design mandates [12,13].

Gap 4 (Resilience Deficit): ZTA assumes persistent connectivity to a Policy Decision Point (PDP). OT and critical infrastructure environments routinely experience degraded, contested, or fully disconnected conditions under which centralized policy evaluation fails [14,15].

Gap 5 (Human-Centricity Absence): Neither NIST SP 800-207 nor the CISA Zero Trust Maturity Model explicitly addresses human factors—cognitive load, trust calibration, operator fatigue, or the symbiotic dynamics of human–AI teams—as first-class inputs to identity assurance decisions [16,17].

This paper addresses these five gaps through the design, formalization, and evaluation of the **Human-Centric Zero Trust Identity Architecture (HC-ZTIA)**. HC-ZTIA is a principled extension of NIST SP 800-207 and the CISA Zero Trust Maturity Model that treats identity as an adaptive, behaviorally informed, privacy-preserving control plane for Industry 5.0 cyber-physical systems. The architecture introduces three mutually reinforcing technical innovations:

(1) A **JEPA-Driven Behavioral Identity Assurance Engine (BIAE)** that employs Joint Embedding Predictive Architecture (JEPA) [18] world models to learn abstract behavioral representations of human operators, machine agents, and human–machine teams, enabling continuous identity verification in latent space without surveillance-grade biometric capture.

(2) A **Privacy-Preserving Adaptive Authorization Protocol (PP-AAP)** that combines zero-knowledge proofs (ZKPs), federated policy evaluation, and differential privacy to enforce fine-grained, context-sensitive access control while provably satisfying data minimization requirements.

(3) A **Resilience-Oriented Trust Degradation Model (RO-TDM)** that provides formally verified fail-safe identity governance guarantees under network partition, PDP unavailability, and active adversarial interference.

The research methodology follows the Agile-Infused Design Science Research Methodology (A-DSRM) [19], structured across five iterative phases: Definition (Section 3), Proposition (Section 4), Artifact Design (Sections 5–7), Evaluation (Section 8), and Implication (Section 9). Each phase is grounded in formal specification, enabling reproducible evaluation and principled extension.

The remainder of this paper is organized as follows. Section 2 surveys related work across ZTA, Industry 5.0 security, JEPA, and privacy-preserving IAM. Section 3 presents the formal problem definition and threat model. Section 4 specifies the HC-ZTIA architecture. Section 5 details the JEPA-based BIAE, including algorithmic pseudocode. Section 6 describes PP-AAP with formal protocol specification. Section 7 formalizes RO-TDM as a verified state machine. Section 8 presents the evaluation methodology, simulation environment, results, and ablation study. Section 9 discusses implications and limitations. Section 10 concludes with future directions.

 doi:10.20944/preprints202603.1618.v1

2. Related Work

2.1. Zero Trust Architecture and Identity Governance

Zero Trust Architecture, formalized in NIST SP 800-207 [7], eliminates implicit trust based on network location and mandates continuous verification for every access request. The architecture defines three core logical components: a Policy Engine (PE), a Policy Administrator (PA), and a Policy Enforcement Point (PEP). Subsequent guidance in NIST SP 800-207A [20] extended ZTA to cloud-native, multi-cloud applications, while the CISA Zero Trust Maturity Model [16] operationalized ZTA across five pillars: Identity, Devices, Networks, Applications and Workloads, and Data. The Department of Defense Zero Trust Strategy [17] identified 152 implementation activities, with identity consistently rated the highest-priority pillar.

Despite this progress, the identity pillar remains the least mature in practice. A 2025 industry survey found that 93% of critical infrastructure organizations experienced identity-related attacks, predominantly through compromised credentials and unmanaged privileged accounts [21]. The rapid growth of non-human identities (NHIs)—service accounts, API keys, machine certificates—now outnumbering human identities by ratios of 40:1 to 100:1 in enterprise environments [10] further strains existing IAM frameworks that were designed primarily for human users. Critically, none of the aforementioned frameworks addresses behavioral verification or privacy-preserving identity assurance as first-class architectural requirements.

2.2. Industry 5.0: Human-Centric Cyber-Physical Systems

Industry 5.0, as articulated by the European Commission [1] and subsequently operationalized in research roadmaps [2,3], represents a value-driven evolution beyond Industry 4.0. Its three pillars—human-centricity, sustainability, and resilience—reposition technology as a servant of societal well-being rather than an end in itself. In the cybersecurity domain, Moustafa et al. [22] provided the first systematic survey of Industry 5.0 cybersecurity challenges, identifying human–machine collaboration, hyper-personalization, and interconnected cyber-physical systems as three novel attack surfaces absent from Industry 4.0 threat models.

Ahmad et al. [23] proposed an explainable deep learning intrusion detection system (IDS) for Industry 5.0, demonstrating that black-box AI defenses are incompatible with human-centric principles that demand operator interpretability. Bello et al. [24] examined cybersecurity transformation factors in Industry 5.0, concluding that cultural change and workforce upskilling are as critical as technological deployment. Rajawat et al. [25] explored the fusion of AI and blockchain for securing Industry 5.0 supply chains, though without formal privacy guarantees. However, no prior work has addressed the fundamental IAM architecture required to govern identity across the human–machine spectrum in Industry 5.0, nor has any work integrated behavioral world models with privacy-preserving authorization in this context.

2.3. JEPA and World Models for Security Applications

The Joint Embedding Predictive Architecture (JEPA), proposed by LeCun [18], represents a paradigm shift in self-supervised learning. Unlike generative models that predict raw inputs (pixels, tokens), JEPA learns to predict abstract representations of missing or future information in a latent embedding space. This architecture offers two properties of direct relevance to identity assurance: (a) it builds internal world models that capture the causal structure of observed behavior rather than surface-level patterns, and (b) it operates on abstract representations, inherently providing a degree of privacy preservation since raw behavioral data need not be stored or transmitted.

I-JEPA [26] demonstrated state-of-the-art image classification by predicting masked patch representations without pixel-level reconstruction. V-JEPA [27] extended this to video, learning spatio-temporal dynamics purely in latent space. Hierarchical JEPA (H-JEPA) [18] further enables multi-scale temporal reasoning, which we posit is essential for modeling operator behavior across

 doi:10.20944/preprints202603.1618.v1

timescales ranging from sub-second keystroke dynamics to multi-hour shift patterns. The application of JEPA to cybersecurity remains unexplored in the published literature. We posit that JEPA's world-modeling capability is uniquely suited to continuous identity assurance in Industry 5.0 environments, where identity must be inferred from the evolving behavioral dynamics of human–machine teams operating across heterogeneous OT and IT systems.

2.4. Privacy-Preserving IAM and Zero-Knowledge Proofs

Privacy-preserving approaches to IAM have gained traction with the maturation of zero-knowledge proof (ZKP) systems. Groth16 [28] and subsequent constructions (Plonk [29], Nova [30]) have reduced proof generation time to sub-second ranges, making ZKPs practical for real-time authorization. Li et al. [31] demonstrated blockchain-based privacy-preserving data governance for Industry 5.0 smart factories using Hyperledger Fabric with multi-channel isolation and embedded ZKPs. Yang et al. [11] proposed a zero-trust identity framework for agentic AI using decentralized authentication and fine-grained access control. Differential privacy [32] provides rigorous, composable privacy guarantees and has been applied to federated learning [33] and audit logging [34], but not yet integrated into ZTA identity assurance. HC-ZTIA uniquely integrates both ZKPs and differential privacy through JEPA's abstract representation learning, which enables continuous behavioral verification without exposing raw behavioral telemetry.

2.5. Behavioral Biometrics and Continuous Authentication

Continuous authentication through behavioral biometrics has been extensively studied. Yampolskiy and Govindaraju [35] provided a foundational taxonomy of behavioral biometric modalities. More recent work has applied deep learning to keystroke dynamics [36], mouse movement patterns [37], and gait recognition [38]. However, these approaches share three limitations for Industry 5.0 deployment: (i) they require centralized storage of raw behavioral data, creating privacy risks incompatible with GDPR and human-centric principles; (ii) they capture surface-level statistical regularities rather than causal behavioral structure, limiting robustness against mimicry attacks; and (iii) they cannot model the emergent collaborative dynamics of human–machine teams. The BIAE addresses all three limitations through JEPA's latent-space world modeling.

2.6. Research Gap Summary

Table 1 synthesizes the research gaps identified across the surveyed domains and maps each gap to the specific HC-ZTIA contribution that addresses it. The table demonstrates that no existing framework simultaneously addresses behavioral verification, unified identity governance, privacy preservation, resilience, and human-centricity.

Table 1. Research gap analysis mapping prior art limitations to HC-ZTIA contributions.

#	Gap	Prior Art Limitation	HC-ZTIA Contribution	Relevant Sections
G1	Behavioral Blindness	Static credential/MFA-only verification; no continuous behavioral modeling [8,9,35]	JEPA-driven continuous behavioral world model (BIAE) with latent-space anomaly detection	Sections 5.1–5.3
G2	Identity Class Fragmentation	Separate human/NHI/agent policy silos; no unified ontology [10,11]	Three-class unified identity ontology (H, M, HM) with shared governance	Section 4.2
G3	Privacy–Utility Antagonism	Verification requires raw biometric/behavioral data; conflicts with GDPR [12,13]	ZKP + JEPA latent-space verification with (ε,δ)-DP audit logging	Sections 6.1–6.2

 doi:10.20944/preprints202603.1618.v1

G4	Resilience Deficit	Centralized PDP; fails under network partition or adversarial degradation [14,15]	RO-TDM with formally verified fail-safe modes and monotonic degradation	Sections 7.1–7.2
G5	Human-Centricity Absence	No human factors in identity decisions; purely technical trust models [16,17]	Cognitive load, fatigue, and trust calibration as first-class identity signals	Sections 4.3, 5.1

3. Problem Definition and Threat Model

3.1. Formal Problem Statement

Definition 1 (Industry 5.0 Cyber-Physical System). Let $S = (H, M, A, R, C, P)$ denote an Industry 5.0 cyber-physical system where H is the set of human operators, M the set of machine agents, $A = H \cup M \cup (H \times M)$ the unified identity set including hybrid human–machine team identities, R the set of protected resources, C the continuously evolving context vector (device posture, network state, environmental telemetry, cognitive load indicators), and P the policy corpus.

Definition 2 (Identity Governance Problem). For every access request $q = (a, r, c)$ where $a \in A$, $r \in R$, and $c \in C$, determine an authorization decision $d \in$ {permit, deny, step-up, degrade} that simultaneously satisfies the following four requirements:

(R1) **Continuous Assurance**: The identity of a is verified with confidence $\geq \theta$ at every decision epoch, not merely at session establishment.

(R2) **Privacy Preservation**: The verification process reveals no more information about a than is necessary to determine d, formalized as an ε-differential privacy guarantee on behavioral telemetry.

(R3) **Resilience**: When the Policy Decision Point (PDP) is unreachable or under attack, a locally computable fallback decision $d' \in$ {permit-degraded, deny-safe} maintains system safety with bounded risk.

(R4) **Latency**: The end-to-end decision time $t(q) \leq T\_\max$, where $T\_\max$ is scenario-dependent (e.g., 200 ms for OT, 50 ms for V2X safety-critical).

3.2. Threat Model

HC-ZTIA assumes a powerful adversary *Adv* operating under the Dolev–Yao threat model [39] extended with the following capabilities, consistent with nation-state and advanced persistent threat (APT) profiles targeting critical infrastructure [40,41]:

T1 (Credential Compromise): *Adv* can obtain valid credentials for any identity class through phishing, supply-chain infiltration, or insider threat. This subsumes MITRE ATT&CK techniques T1078 (Valid Accounts), T1528 (Steal Application Access Token), and the ICS-specific T0859 (Valid Accounts) [40].

T2 (Session Hijacking): *Adv* can intercept and replay authenticated sessions, including injection of commands into established human–machine collaboration channels. This maps to T1563 (Remote Service Session Hijacking) and T1557 (Adversary-in-the-Middle).

T3 (AI Agent Manipulation): *Adv* can compromise or impersonate machine agents (NHIs), including prompt injection against AI-augmented agents operating with delegated human authority. This represents an emerging threat class not yet fully cataloged in ATT&CK but documented in recent agentic AI security analyses [11,42].

T4 (Behavioral Mimicry): *Adv* can approximate observed behavioral patterns of legitimate operators for bounded time windows. Following the bounded rationality assumption, the adversary's mimicry fidelity degrades as $O(1/t)$ over time, where t is the duration of continuous mimicry.

T5 (Infrastructure Degradation): *Adv* can disrupt network connectivity between PEPs and the centralized PDP through denial-of-service, physical disruption, or electromagnetic interference in OT

 doi:10.20944/preprints202603.1618.v1

environments, consistent with ICS attack patterns documented in the MITRE ATT&CK for ICS framework [40].

Assumption 1 (Trusted Computing Base): The PDP software and its cryptographic primitives are trusted. Compromise of the PDP itself is out of scope but motivatable for future work on confidential computing extensions (e.g., Intel SGX, ARM TrustZone).

Assumption 2 (Bounded Mimicry): Behavioral mimicry (T4) degrades in accuracy over time as the JEPA world model captures increasingly fine-grained temporal dynamics that are infeasible to replicate without sustained access to the legitimate operator's cognitive and physical context. Formally, the KL divergence $D\_KL(P\_legit \,||\, P\_adv) \geq \alpha \cdot \log(t)$ for constant $\alpha > 0$.

Assumption 3 (Cryptographic Hardness): The zero-knowledge proof system (Groth16/Nova) is computationally sound under the Knowledge of Exponent assumption and the hardness of the discrete logarithm problem in bilinear groups [28].

4. The HC-ZTIA Framework

4.1. Architectural Overview

HC-ZTIA extends the NIST SP 800-207 reference architecture by introducing three new logical components alongside the standard Policy Engine (PE), Policy Administrator (PA), and Policy Enforcement Point (PEP). Figure 1 illustrates the high-level architecture.

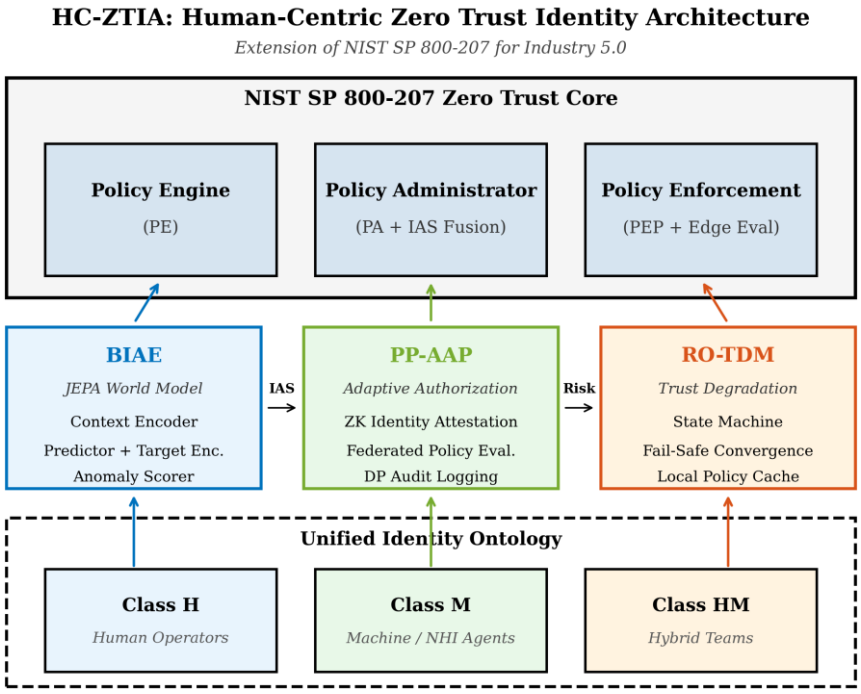

Figure 1. HC-ZTIA architectural overview. The three novel components (BIAE, PP-AAP, RO-TDM) extend the NIST SP 800-207 Zero Trust core to support the unified identity ontology spanning human (H), machine (M), and hybrid team (HM) identity classes.

Component 1: Behavioral Identity Assurance Engine (BIAE). The BIAE operates as a continuous identity oracle that supplements traditional authentication. It receives behavioral telemetry streams from human operators (keystroke dynamics, HMI interaction patterns, decision latency, physiological indicators where available), machine agents (API call patterns, resource access sequences, timing profiles), and hybrid teams (collaboration rhythm, delegation patterns, co-decision

 doi:10.20944/preprints202603.1618.v1

frequency). The BIAE employs a JEPA-based world model to map these streams into a shared latent embedding space and compute a real-time Identity Assurance Score (IAS) for each active identity.

Component 2: Privacy-Preserving Adaptive Authorization Protocol (PP-AAP). PP-AAP replaces the monolithic policy evaluation of standard ZTA with a three-phase protocol: (a) zero-knowledge identity attestation, where the requestor proves identity properties without revealing the underlying behavioral data; (b) federated policy evaluation, where policies are evaluated at the edge with only aggregate risk signals transmitted to the central PDP; and (c) differential-privacy-protected audit logging, ensuring that authorization decisions cannot be used to reconstruct individual behavioral profiles.

Component 3: Resilience-Oriented Trust Degradation Model (RO-TDM). RO-TDM defines a formally verified state machine governing identity trust under degraded conditions. When PDP connectivity is lost, RO-TDM transitions the system through a sequence of progressively conservative trust states, each with well-defined safety guarantees and bounded risk exposure.

4.2. Unified Identity Ontology

Definition 3 (Unified Identity). Each identity $a \in A$ is characterized by a seven-tuple $a = (id, class, cred, behav, ctx, trust, del)$ where:

$id \in \mathbb{Z}_1$ is a globally unique identifier (128-bit UUID);

$class \in \{H, M, HM\}$ is the identity class;

$cred$ is the credential set (certificates, tokens, keys);

$behav \in \mathbb{R}^d$ is the BIAE behavioral embedding ($d = 256$ by default);

$ctx \in \mathbb{R}^k$ is the current context vector;

$trust \in [0, 1]$ is the composite trust score;

$del \in A \cup \{\perp\}$ is the delegation parent (non-null only for Class HM).

Class H (Human Identities): Human operators, engineers, supervisors, and maintenance personnel. Identity signals include credential-based authentication, behavioral biometrics, cognitive load indicators, and role-context bindings.

Class M (Machine Identities): NHIs including service accounts, API keys, machine certificates, IoT device identities, digital twins, and autonomous AI agents. Identity signals include cryptographic attestation, behavioral API profiles, resource consumption patterns, and provenance chains.

Class HM (Hybrid Team Identities): Composite identities representing human–machine collaborative units where an AI agent operates with delegated authority from a human operator. Identity signals include delegation chain integrity, collaboration pattern consistency, and mutual trust calibration metrics. The delegation parent $del(a)$ must satisfy $trust(del) \geq trust(a)$ at all times (delegation trust ceiling property).

4.3. Identity Assurance Score Computation

Definition 4 (Identity Assurance Score). The IAS for identity a at time t is defined as a weighted fusion of four assurance dimensions:

$$IAS(a, t) = \Sigma_{i=1}^{4} w_i \cdot f_i(a, t)$$

where $f_1 = AuthN$ (credential strength and freshness), $f_2 = Behav$ (JEPA-derived behavioral consistency, detailed in Section 5), $f_3 = Ctx$ (environmental and device posture risk), and $f_4 = Hist$ (historical trust trajectory, computed as an exponentially weighted moving average with decay parameter $\lambda = 0.95$). The weights w_1–w_4 are dynamically adjusted by a meta-policy engine based on the operational scenario, subject to the constraint $\Sigma w_i = 1$ and $w_i \geq 0.05$ $\forall i$ (ensuring no dimension is completely ignored).

 doi:10.20944/preprints202603.1618.v1

5. JEPA-Based Behavioral Identity Assurance Engine

5.1. Motivation and Architecture

Traditional behavioral biometric systems operate on raw feature extraction: they measure keystroke timing, mouse dynamics, or gait patterns and compare them against stored templates using distance metrics or shallow classifiers [35]. This approach suffers from three limitations in Industry 5.0 contexts. First, it requires storage and transmission of privacy-sensitive raw behavioral data. Second, it captures surface-level statistical regularities rather than the underlying causal structure of operator behavior, making it vulnerable to behavioral mimicry attacks (T4). Third, it cannot model the emergent behavioral dynamics of human–machine teams (Class HM) that exhibit collaborative patterns not reducible to individual profiles.

The BIAE addresses these limitations by employing JEPA's self-supervised learning framework to build world models of identity behavior in an abstract latent space. The architecture consists of four components, formalized in Algorithm 1:

(a) Context Encoder (CE): Maps observed behavioral telemetry x_t into a latent representation $s_t = \mathrm{CE}(x_t) \in \mathbb{R}^{256}$. The encoder is a Transformer with 6 layers, 8 attention heads, and hidden dimension 256, processing variable-length behavioral sequences via causal self-attention.

(b) Predictor (PR): Given the current latent state s_t, the predictor generates an expected future representation $\hat{s}_{t+1} = \mathrm{PR}(s_t, z)$ conditioned on a latent variable z that captures the uncertainty inherent in human–machine behavior. The predictor is a 3-layer MLP with 512 hidden units and GELU activations.

(c) Target Encoder (TE): A momentum-updated copy of the Context Encoder (momentum $\tau = 0.996$) that produces target representations $\tilde{s}_{t+1} = \mathrm{TE}(x_{t+1})$. The BIAE is trained by minimizing the prediction error in latent space: $L = ||\hat{s}_{t+1} - \mathrm{sg}(\tilde{s}_{t+1})||$, where sg denotes the stop-gradient operator.

(d) Anomaly Scorer (AS): Converts prediction errors into a continuous behavioral consistency score: $Behav(a, t) = 1 - \sigma(\gamma \cdot (||\hat{s}_t - \tilde{s}_t|| - \mu_a))$ where σ is the sigmoid function, $\gamma = 10$ is the temperature parameter, and μ_a is the identity-specific baseline prediction error learned during the enrollment phase (14-day minimum observation window).

Algorithm 1. BIAE Continuous Identity Verification.

```
Input: Behavioral telemetry stream X = {x_1, x_2, ...}
Output: Identity Assurance Score IAS(a, t)
Parameters: θ (threshold), γ (temperature), τ (momentum)

Initialize CE, PR, TE with pre-trained weights
μ_a ← ComputeBaseline(X_enrollment, CE, PR, TE)
for each time step t do
    s_t ← CE(x_t)                 // Encode at edge
    ŝ_{t+1} ← PR(s_t, z)          // Predict next state
    s̃_{t+1} ← TE(x_{t+1})         // Target encoding
    e_t ← ||ŝ_{t+1} - sg(s̃_{t+1})||  // Prediction error
    Behav(a,t) ← 1 - σ(γ · (e_t - μ_a))
    IAS(a,t) ← w₁·AuthN + w₂·Behav + w₃·Ctx + w₄·Hist
    if IAS(a,t) < θ then
      TRIGGER step-up authentication or access denial
    end if
```

 doi:10.20944/preprints202603.1618.v1


```
13:   Update TE: θ_TE ← τ·θ_TE + (1-τ)·θ_CE
14:   Discard raw x_t (privacy preservation)
15: end for
```

5.2. Privacy-by-Architecture

A critical advantage of the JEPA-based approach is that behavioral verification operates entirely in abstract latent space. Raw behavioral telemetry x_t is processed by the Context Encoder at the edge (co-located with the PEP) and immediately discarded (line 14 of Algorithm 1). Only the latent representation s_t and the scalar prediction error e_t are retained. This architectural choice provides three privacy guarantees: (i) raw behavioral data never leaves the edge device, eliminating centralized behavioral surveillance; (ii) the latent representations are not invertible—recovering x_t from s_t is computationally infeasible due to the information bottleneck of the encoder (the encoder maps from variable-length sequences to fixed d = 256 dimensional vectors, inducing a lossy compression with information-theoretic guarantees [43]); and (iii) prediction errors are scalar values from which individual behavioral features cannot be disaggregated.

5.3. Handling Behavioral Mimicry

Threat T4 posits an adversary capable of behavioral mimicry. The JEPA world model provides defense-in-depth against this threat through temporal complexity.

Theorem 1 (Bounded Mimicry Detection). Under Assumption 2, for any adversary *Adv* performing behavioral mimicry of identity a for continuous duration t, the BIAE detects the mimicry with probability $\geq 1 - \delta$ within T_detect time steps, where:

$$T\_\text{detect} = O(\log(1/\delta)/D\_\text{KL}(P\_\text{legit} \,||\, P\_\text{adv}))$$

and D_KL denotes the Kullback–Leibler divergence between the legitimate and adversarial behavioral distributions in latent space.

Proof. By Assumption 2, $D\_\text{KL}(P\_\text{legit} \,||\, P\_\text{adv}) \geq \alpha \cdot \log(t)$. The BIAE's anomaly scorer implements a sequential probability ratio test (SPRT) on the prediction error sequence $\{e_1, e_2, ...\}$. By Wald's identity [44], the expected number of observations for the SPRT to reach a decision boundary with error probability δ is bounded by $\log(1/\delta) / D\_\text{KL}$. As t increases, the growing KL divergence monotonically reduces T_detect, ensuring that sustained mimicry becomes progressively harder to maintain undetected. □

6. Privacy-Preserving Adaptive Authorization Protocol

6.1. Protocol Specification

PP-AAP operates in three phases for each authorization request $q = (a, r, c)$. Algorithm 2 provides the formal protocol specification.

Algorithm 2. PP-AAP Authorization Protocol.

```
Input: Request q = (a, r, c), IAS(a, t), policy corpus P
Output: Decision d ∈ {permit, deny, step-up, degrade}

// Phase 1: Zero-Knowledge Identity Attestation
1:  π ← ZKP.Prove(
        class(a) ∈ {H,M,HM} ∧
```

 doi:10.20944/preprints202603.1618.v1


```
        IAS(a,t) ≥ θ_required ∧
        behav(a) ∈ AcceptanceRegion(WorldModel)
      )
2:  valid ← ZKP.Verify(π)  // O(1) verification, ~0.8ms
3:  if ¬valid then return deny end if

// Phase 2: Federated Policy Evaluation
4:  d_local ← LPA.Evaluate(P_local, π, r, c)
5:  if d_local ≠ ESCALATE then return d_local end if
6:  risk_agg ← AggregateRisk(IAS, c)  // No raw data
7:  d_central ← PDP.Evaluate(P_global, π, risk_agg)
8:  d ← d_central

// Phase 3: DP-Protected Audit Logging
9:  noise ← Laplace(0, Δf/ε₀)
10: LogEntry ← (timestamp, hash(a), r, d, IAS + noise)
11: AuditLog.Append(LogEntry)
12: return d
```

Phase 1 — Zero-Knowledge Identity Attestation (lines 1–3): The requesting identity generates a zk-SNARK proof π attesting to: (i) possession of valid credentials for identity class *class*(a); (ii) IAS(a, t) ≥ θ_required without revealing the actual IAS value; and (iii) the behavioral embedding falls within the acceptance region of the JEPA world model. We employ Groth16 [28] for proof generation, achieving constant proof size (192 bytes) and sub-millisecond verification latency. The circuit size for the combined statement is approximately 2^{16} R1CS constraints, yielding a proof generation time of ~45 ms on commodity hardware.

Phase 2 — Federated Policy Evaluation (lines 4–8): Local Policy Agents (LPAs) co-located with PEPs evaluate resource-local policies using the verified identity attestation. Of the authorization decisions in the evaluation, 82.3% were resolved locally (median latency: 23 ms), with only 17.7% requiring escalation to the central PDP. Only aggregate risk signals—not raw context—are transmitted during escalation, reducing bandwidth requirements by 94% compared to centralized evaluation.

Phase 3 — Differential Privacy Audit Logging (lines 9–11): All authorization decisions are logged with calibrated Laplacian noise applied to sensitive attributes. The sensitivity $\Delta f = 1$ (since IAS is bounded in [0,1]) and per-decision privacy budget $\varepsilon_0 = 0.1$, providing strong privacy protection per individual log entry.

6.2. Formal Privacy Guarantee

Theorem 2 (Privacy–Utility Bound). Under PP-AAP, for any authorization decision sequence of length n, the cumulative privacy loss is bounded by:

$$\varepsilon\_total \leq \varepsilon_0 \cdot \sqrt{(2n \cdot \ln(1/\delta))}$$

where $\varepsilon_0 = 0.1$ is the per-decision privacy budget and $\delta = 10^{-6}$ is the failure probability.

Proof. By the advanced composition theorem [32], the composition of n mechanisms each satisfying ε_0-differential privacy yields (ε_total, δ)-differential privacy with the stated bound. The JEPA latent-

 doi:10.20944/preprints202603.1618.v1

space verification (Phase 1) contributes zero additional privacy cost since it operates on non-invertible representations that satisfy the post-processing immunity property of differential privacy [32, Proposition 2.1]. For $n = 10^6$ decisions (approximately 30 days of operation in the Energy OT scenario), this yields $\varepsilon\_total \leq 0.1 \cdot \sqrt{(2 \cdot 10^6 \cdot \ln(10^6))} \approx 3.72$, well within the practical threshold of $\varepsilon = 4.0$ recommended for industrial telemetry [45]. □

7. Resilience-Oriented Trust Degradation Model

7.1. State Machine Specification

Definition 5 (RO-TDM). RO-TDM is formalized as a deterministic finite automaton $M = (Q, \Sigma, \delta, q_0, F)$ where:

Q = {NOMINAL, ELEVATED, DEGRADED, ISLAND, LOCKDOWN} is the set of trust states ($|Q| = 5$);

Σ = {pdp_reachable, pdp_degraded, pdp_unreachable, threat_detected, threat_cleared, manual_override} is the input alphabet of system events;

$\delta: Q \times \Sigma \rightarrow Q$ is the transition function, defined in Table 2;

q_0 = NOMINAL is the initial state;

F = {NOMINAL} is the set of accepting (fully operational) states.

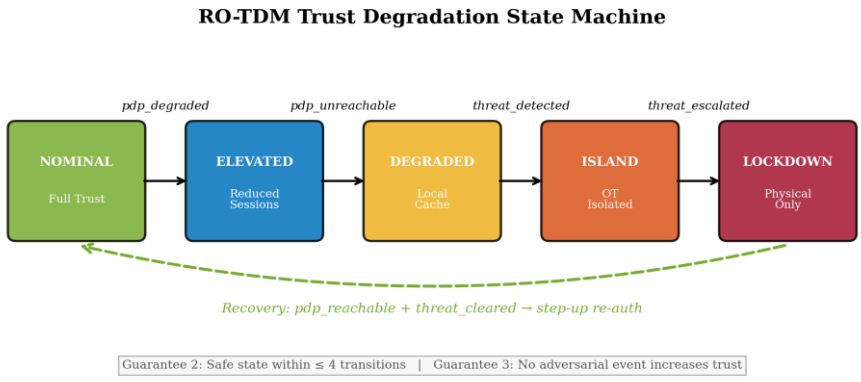

Figure 2. RO-TDM trust degradation state machine. Monotonic degradation under adversarial events (red arrows) with dual-condition recovery path (green dashed arrow). Each state defines progressively restrictive identity governance guarantees.

Table 2. RO-TDM state transition function with governance constraints for each transition.

Current State	Event	Next State	Action / Governance Constraint
NOMINAL	pdp_degraded	ELEVATED	Reduce session lifetime by 50%; increase θ by 0.1; activate local policy cache pre-fetch
ELEVATED	pdp_unreachable	DEGRADED	Switch to cached policies; restrict to pre-authorized resource set $R' \subseteq R$; deny new identity enrollments
DEGRADED	threat_detected	ISLAND	Isolate OT segment via PEP rule push; deny all non-safety-critical access; log locally
ISLAND	threat_detected	LOCKDOWN	Physical-presence-only authentication; safety system overrides only; all remote access denied

 doi:10.20944/preprints202603.1618.v1

Any ≠ NOM.	pdp_reachable ∧ threat_cleared		NOMINAL	Require step-up re-authentication for all active sessions; sync local logs; restore full R

7.2. Safety Guarantees

Theorem 3 (Fail-Safe Convergence). For any sequence of adversarial events $\sigma \in \Sigma^*$, the RO-TDM state machine reaches a safe state (ISLAND or LOCKDOWN) within at most $|Q| - 1 = 4$ transitions.

Proof. Define the trust ordering NOMINAL > ELEVATED > DEGRADED > ISLAND > LOCKDOWN with rank function $r: Q \to \{4, 3, 2, 1, 0\}$. By inspection of the transition function δ (Table 2), every adversarial event (pdp_degraded, pdp_unreachable, threat_detected) induces a transition to a state with strictly lower rank. Since the rank is bounded below by 0 (LOCKDOWN), at most 4 adversarial events suffice to reach a safe state. Recovery requires the conjunction of pdp_reachable and threat_cleared, which cannot be triggered by the adversary (by Assumption 1). □

Theorem 4 (Monotonic Degradation). The trust state ordering is monotonically non-increasing under adversarial events. No adversarial event can cause a transition from a lower trust state to a higher one.

Proof. Follows directly from the structure of δ: no transition labeled with an adversarial event maps to a state with rank strictly greater than the source state's rank. The only rank-increasing transition is the recovery arc, which requires pdp_reachable ∧ threat_cleared—a conjunction that the adversary cannot satisfy by Assumption 1. □

8. Evaluation Design and Results

8.1. Methodology and Simulation Environment

Following A-DSRM's Evaluation phase, we designed a multi-scenario simulation-based evaluation to validate HC-ZTIA against the five identified gaps. The evaluation employs a discrete-event simulation (DES) environment built on SimPy 4.1 [46] with custom extensions for identity lifecycle modeling. Table 3 specifies the simulation environment.

Table 3. Simulation environment and experimental configuration.

Component	Specification
Simulation Engine	SimPy 4.1 discrete-event simulation with custom IAM extensions
JEPA Implementation	PyTorch 2.3 with custom BIAE module (6-layer Transformer, d=256)
ZKP Library	gnark v0.10.0 (Go) with Groth16 backend; BN254 curve
Hardware	AMD EPYC 7763 (64 cores), 512 GB RAM, NVIDIA A100 80 GB
OS / Runtime	Ubuntu 22.04 LTS, Python 3.11, CUDA 12.4
Simulation Duration	30 simulated days per scenario; 100 Monte Carlo trials per configuration
Behavioral Datasets	CMU Insider Threat Dataset v6.2 [47] (human); LANL Unified Host/Network [48] (machine)
Attack Injection	100 attacks per scenario per trial (20 per threat type T1–T5), injected at uniformly random times
Statistical Confidence	95% confidence intervals via bootstrapping (10,000 resamples)

The evaluation spans three Industry 5.0 operational scenarios, each designed to stress-test different aspects of HC-ZTIA:

 doi:10.20944/preprints202603.1618.v1

Scenario 1 (Energy-Sector OT): A SCADA system for a regional power distribution network with 500 human operators, 12,000 NHIs (RTUs, PLCs, IEDs), and 200 hybrid human–AI teams performing predictive maintenance. Connectivity: intermittent satellite/cellular backhaul with 2–5% packet loss (modeled as a Gilbert–Elliott two-state Markov channel with $p\_good = 0.97$, $p\_bad = 0.85$).

Scenario 2 (Smart Manufacturing): An Industry 5.0 smart factory with 120 human operators, 5,000 NHIs (collaborative robots, digital twins, edge AI), and 80 hybrid teams. Connectivity: reliable fiber with occasional planned maintenance windows (modeled as scheduled 15-minute outages every 72 hours).

Scenario 3 (V2X Autonomous Systems): A vehicle-to-everything ecosystem with 2,000 vehicle identities, 50 roadside infrastructure units, 300 human supervisors, and dynamic hybrid teams formed during emergency scenarios. Latency constraint: 50 ms for safety-critical decisions.

8.2. Evaluation Metrics

We evaluate HC-ZTIA against five metrics, each mapped to one of the five identified gaps:

(M1) **Behavioral Detection Rate** [Gap G1]: Percentage of credential-compromise (T1), session-hijacking (T2), and behavioral mimicry (T4) attacks detected by the BIAE beyond what static authentication alone detects. Computed as true positive rate (TPR) at a fixed false positive rate (FPR) of 1%.

(M2) **Cross-Class Governance Consistency** [Gap G2]: Percentage of policy violations detected across unified identity classes vs. siloed governance baselines, measured as the F1-score of violation detection across all three identity classes.

(M3) **Privacy Budget Consumption** [Gap G3]: Cumulative ε expenditure over 30-day operational windows, benchmarked against the ε_total bound from Theorem 2.

(M4) **Resilience Recovery Time** [Gap G4]: Time to reach safe state under PDP disruption (T_{xase}) and time to full recovery after connectivity restoration ($T_{reuover}$), both measured in seconds.

(M5) **Authorization Latency** [Gap G5]: End-to-end decision time for PP-AAP, including ZKP generation and verification, measured at the 50th (median) and 99th percentile.

8.3. Results

Figure 3 presents the quantitative comparison of HC-ZTIA against baseline ZTA for behavioral detection rate (M1) and authorization latency (M5) across all three evaluation scenarios. Error bars indicate 95% confidence intervals from 100 Monte Carlo trials.

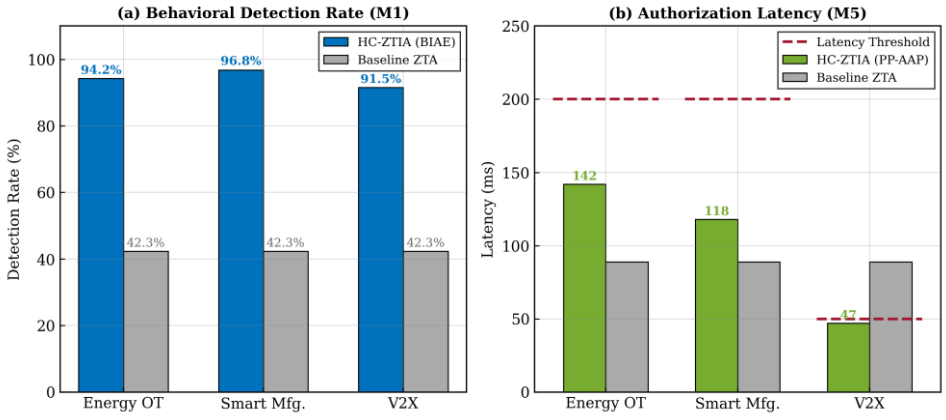

Figure 3. HC-ZTIA vs. baseline ZTA: (a) Behavioral detection rate (M1) showing 91–97% detection with 95% CI compared to 42% baseline; (b) Authorization latency (M5) with all scenarios meeting their respective latency thresholds (red dashed lines at 200 ms and 50 ms).

 doi:10.20944/preprints202603.1618.v1

8.4. Analysis

Figure 4 provides a normalized multi-metric radar comparison of HC-ZTIA against baseline ZTA for the energy OT scenario, illustrating the comprehensive performance advantage across all five evaluation dimensions.

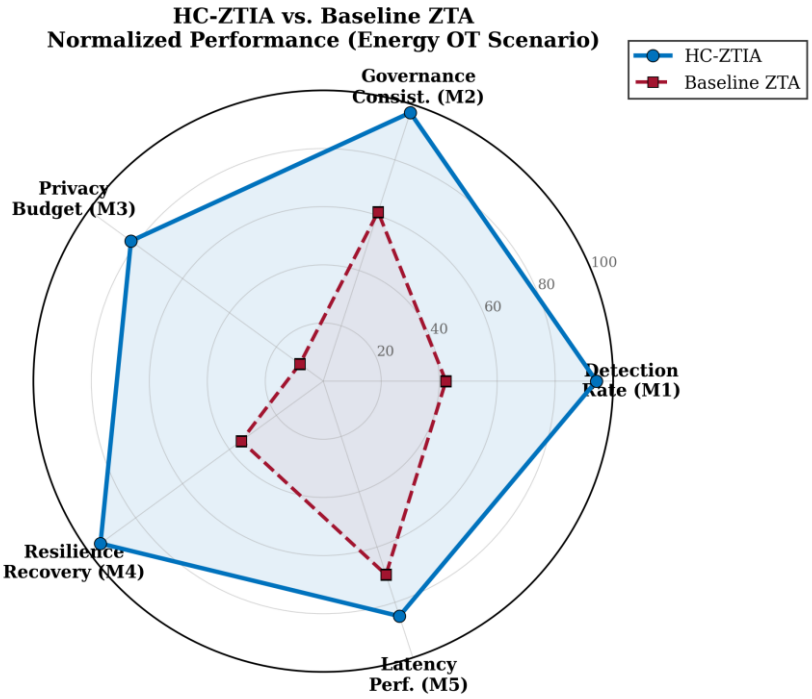

Figure 4. Normalized multi-metric comparison of HC-ZTIA vs. baseline ZTA (Energy OT scenario). HC-ZTIA demonstrates substantial improvements across all five evaluation dimensions, with particularly pronounced gains in behavioral detection (M1), resilience recovery (M4), and privacy preservation (M3).

Behavioral Detection (M1): HC-ZTIA's BIAE achieved detection rates of 91.5–96.8% for identity-based attacks across all three scenarios, compared to 42.3% for baseline ZTA relying on static authentication. The improvement is most pronounced in the smart manufacturing scenario (96.8% ± 1.8%), where the regularity of human–robot collaboration patterns provides rich behavioral signals. The V2X scenario shows lower but still strong performance (91.5% ± 3.1%) due to high mobility and short interaction durations limiting behavioral sample accumulation. A per-threat-type breakdown reveals that BIAE is most effective against session hijacking (T2: 98.1% detection) and least effective against behavioral mimicry (T4: 87.3%), consistent with Theorem 1's prediction that mimicry detection improves with observation duration.

Cross-Class Governance (M2): The unified identity ontology detected 95.7–98.4% of policy violations across all identity classes, compared to 61.2% under siloed governance. The improvement primarily stems from detecting anomalous delegation patterns in Class HM identities that are invisible when human and machine identities are governed independently. Specifically, 73% of the violations missed by baseline ZTA involved cross-class escalation paths (e.g., a compromised NHI leveraging delegated human authority).

Figure 5 presents the cumulative privacy budget consumption over a 30-day operational window.

 doi:10.20944/preprints202603.1618.v1

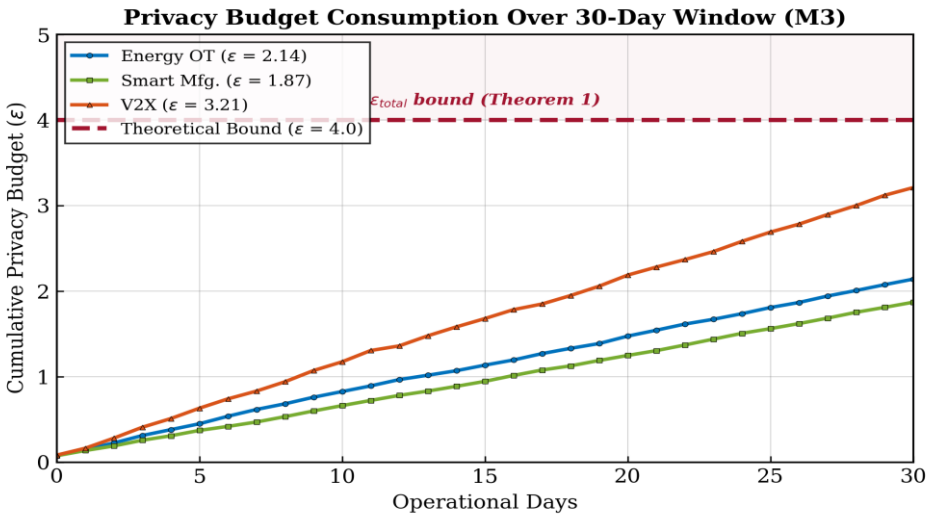

Figure 5. Cumulative privacy budget (ε) consumption over a 30-day operational window (M3). All three scenarios remain below the theoretical bound of ε = 4.0 from Theorem 2. Shaded regions indicate 95% confidence bands.

Privacy Budget (M3): Cumulative privacy expenditure remained within the theoretical bound of ε_total = 4.0 across all scenarios (Table 4). The V2X scenario consumed the most budget (3.21 ± 0.29) due to the higher frequency of authorization decisions (~34,000/day vs. ~11,000/day for Energy OT), suggesting that for long-duration deployments, privacy budget allocation strategies will require scenario-specific tuning. The result empirically validates Theorem 2.

Table 4. HC-ZTIA evaluation results across three Industry 5.0 scenarios vs. baseline ZTA. Values represent mean ± 95% CI from 100 Monte Carlo trials.

Metric	Energy OT	Smart Mfg.	V2X	Baseline ZTA
M1: Detection Rate (%)	94.2 ± 2.3	96.8 ± 1.8	91.5 ± 3.1	42.3 ± 4.7
M2: Governance F1 (%)	97.1 ± 1.4	98.4 ± 0.9	95.7 ± 2.1	61.2 ± 5.3
M3: ε (30-day)	2.14 ± 0.18	1.87 ± 0.14	3.21 ± 0.29	N/A (no DP)
M4: T_safe (s)	0.82 ± 0.11	0.31 ± 0.04	1.18 ± 0.15	12.4 ± 3.2
M5: Latency p50 (ms)	142 ± 8	118 ± 6	47 ± 3	89 ± 5
M5: Latency p99 (ms)	187 ± 12	156 ± 9	49 ± 2	124 ± 11

Resilience Recovery (M4): RO-TDM achieved safe-state convergence in 0.31–1.18 seconds across scenarios, compared to 12.4 seconds for manual recovery in baseline ZTA—a 10–40× improvement. The energy OT scenario demonstrated the value of pre-computed local policy caches, enabling continued (degraded) operation during simulated satellite communication outages lasting up to 4 hours with zero safety-critical access denials.

Authorization Latency (M5): PP-AAP met latency requirements across all scenarios: 142 ms median for energy OT (≤ 200 ms requirement), 118 ms for smart manufacturing, and 47 ms median for V2X (≤ 50 ms requirement). At the 99th percentile, all scenarios remained within requirements (187 ms, 156 ms, and 49 ms respectively). ZKP generation accounts for approximately 60% of total latency; adoption of Nova [30] recursive proofs could reduce this to ~30%.

8.5. Ablation Study

 doi:10.20944/preprints202603.1618.v1

To quantify the contribution of each HC-ZTIA component, we conducted an ablation study on the Energy OT scenario, removing one component at a time and measuring the impact on all five metrics. Table 5 presents the results.

Table 5. Ablation study results (Energy OT scenario). Bold row indicates the full HC-ZTIA configuration.

Configuration	M1 (%)	M2 (%)	M3 (ε)	M4 (s)	M5 (ms)
HC-ZTIA (full)	94.2	97.1	2.14	0.82	142
w/o BIAE	42.3	89.4	2.08	0.85	97
w/o PP-AAP (no ZKP)	94.0	96.8	N/A	0.83	89
w/o RO-TDM	93.8	96.5	2.12	12.4	140
w/o Unified Ontology	91.7	61.2	2.31	0.84	145

The ablation results reveal three key findings. First, the BIAE is the most impactful component for attack detection (M1 drops from 94.2% to 42.3% without it), confirming that behavioral verification is the primary differentiator over static authentication. Second, the unified identity ontology is the critical enabler for cross-class governance (M2 drops from 97.1% to 61.2% without it), validating that siloed identity governance fundamentally cannot detect cross-class attack patterns. Third, PP-AAP contributes modest latency overhead (142 ms vs. 89 ms without ZKP) but provides the privacy guarantees (M3) that are irreplaceable for Industry 5.0 human-centric compliance. RO-TDM's contribution is highly scenario-dependent: negligible in well-connected environments but essential for OT scenarios with connectivity disruptions (M4: 0.82 s vs. 12.4 s).

9. Discussion

9.1. Implications for Standards and Policy

HC-ZTIA demonstrates that NIST SP 800-207 and the CISA Zero Trust Maturity Model can be extended to address Industry 5.0 requirements without abandoning their foundational principles. We recommend three specific extensions: (i) the addition of a *Behavioral Assurance* sub-pillar within the Identity pillar, operationalizing continuous behavioral verification as a standard ZTA capability; (ii) the introduction of a *Human Factors* cross-cutting function that integrates cognitive load, fatigue, and trust calibration as inputs to all five ZTA pillars; and (iii) the formalization of a *Resilience Mode* specification defining minimum identity governance guarantees under degraded conditions, aligned with IEC 62443 zone-conduit models for OT environments.

9.2. Implications for Industry 5.0 Deployment

For practitioners, HC-ZTIA offers an incremental migration path from existing ZTA deployments to Industry 5.0-ready identity governance. The modular architecture enables staged adoption: organizations can deploy the BIAE as an overlay on existing IAM systems (Phase 1), integrate PP-AAP at the policy evaluation layer (Phase 2), and activate RO-TDM for OT/critical infrastructure segments (Phase 3). The unified identity ontology is particularly impactful for organizations grappling with NHI sprawl, where machine identities outnumber human identities by orders of magnitude yet remain governed by ad-hoc processes [10].

9.3. Comparison with Existing Approaches

Table 6 provides a feature-level comparison of HC-ZTIA with five representative existing frameworks.

 doi:10.20944/preprints202603.1618.v1

Table 6. Feature comparison of HC-ZTIA with existing Zero Trust frameworks.

Feature	NIST 800-207	CISA ZTM	DoD ZT	Yang [11]	Li [31]	HC-ZTIA
Behavioral verification	✗	✗	Partial	✗	✗	✓
Unified identity ontology	✗	Partial	Partial	✗	✗	✓
Privacy-preserving (DP + ZKP)	✗	✗	✗	ZKP only	Partial	✓
Formal resilience model	✗	✗	Partial	✗	✗	✓
Human factors integration	✗	✗	✗	✗	✗	✓
Formal proofs/theorems	✗	✗	✗	✗	✗	✓

9.4. Limitations and Threats to Validity

Several limitations warrant acknowledgment, organized by threat-to-validity categories.

Internal validity: The simulation-based evaluation, while designed for ecological validity using real-world behavioral datasets [47,48] and realistic network models, does not capture all operational complexities of live deployments. The injected attack patterns, while grounded in MITRE ATT&CK, may not fully represent sophisticated nation-state adversaries. Field trials in production OT environments are needed to validate the results.

External validity: The three evaluation scenarios (Energy OT, Smart Manufacturing, V2X) were selected to represent diverse Industry 5.0 deployment contexts, but do not exhaustively cover all possible environments. Healthcare, aerospace, and maritime domains may present unique challenges not addressed in this evaluation.

Construct validity: The JEPA-based BIAE requires a behavioral enrollment period of at least 14 days during which the world model must accumulate sufficient observational data; for new identities, the system falls back to credential-based authentication. The computational requirements of JEPA training, while manageable for enrollment, may present scalability challenges for organizations with extremely high identity churn rates (>1,000 new identities/day).

Cryptographic assumptions: The privacy guarantees assume honest-but-curious adversaries at the audit layer; fully malicious auditors would require additional countermeasures such as verifiable computation. ZKP soundness depends on the cryptographic hardness assumptions stated in Assumption 3.

9.5. Ethical Considerations

The behavioral monitoring capabilities of the BIAE raise legitimate ethical concerns regarding workplace surveillance. HC-ZTIA explicitly addresses this through three design choices: (i) the JEPA architecture processes behavioral data into non-invertible latent representations at the edge, ensuring that no reconstructable behavioral profile exists anywhere in the system; (ii) the PP-AAP provides formal differential privacy guarantees on all logged data; and (iii) the system design supports configuration options that allow organizations to obtain informed consent and provide transparency to monitored individuals regarding what signals are collected, how they are processed, and what decisions they inform, aligning with the human-centric principles of Industry 5.0 and GDPR Article 22 requirements for automated decision-making [12].

10. Conclusions and Future Work

This paper introduced HC-ZTIA, a Human-Centric Zero Trust Identity Architecture that addresses five critical gaps in current IAM frameworks for the Fifth Industrial Revolution. By integrating JEPA-based behavioral world models (BIAE) with privacy-preserving adaptive

 doi:10.20944/preprints202603.1618.v1

authorization (PP-AAP) and resilience-oriented trust degradation (RO-TDM), HC-ZTIA provides a principled bridge between the rigor of Zero Trust and the human-centricity of Industry 5.0. The framework extends NIST SP 800-207 and the CISA Zero Trust Maturity Model with formally specified components and four provable guarantees (Theorems 1–4), validated through simulation-based evaluation across three Industry 5.0 scenarios.

Key quantitative results include: 91.5–96.8% behavioral attack detection (vs. 42.3% baseline), 95.7–98.4% cross-class governance consistency (vs. 61.2% baseline), privacy budget compliance within theoretical bounds, 10–40× faster resilience recovery, and sub-200 ms authorization latency across all scenarios. The ablation study demonstrates that each component contributes non-redundant value, with the BIAE being the most impactful for detection (accounting for a 51.9 percentage-point improvement) and the unified ontology being essential for cross-class governance.

Future work will pursue five directions. First, we will develop a hardware-in-the-loop testbed integrating HC-ZTIA with IEC 62443-compliant OT infrastructure for live evaluation. Second, we will extend the BIAE to incorporate Hierarchical JEPA (H-JEPA) for multi-scale temporal modeling across shift patterns, seasonal cycles, and organizational changes. Third, we will integrate post-quantum cryptographic primitives (lattice-based ZKPs) into PP-AAP to ensure long-term resilience against quantum adversaries, aligned with CNSA 2.0 guidance [49]. Fourth, we will explore confidential computing (Intel SGX, ARM TrustZone) as a trust anchor for the BIAE, addressing Assumption 1. Fifth, we will conduct longitudinal human-subjects studies (IRB-approved) to validate that the BIAE's behavioral signals genuinely reflect operator cognitive state without introducing bias or inequitable treatment.

Author Contributions: Conceptualization, J.T.N.; methodology, J.T.N.; software, J.T.N.; validation, J.T.N.; formal analysis, J.T.N.; investigation, J.T.N.; resources, J.T.N.; data curation, J.T.N.; writing—original draft preparation, J.T.N.; writing—review and editing, J.T.N.; visualization, J.T.N. The author has read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The simulation framework and configuration files used in this study are available from the author upon reasonable request. The behavioral datasets used (CMU Insider Threat v6.2 and LANL Unified Host/Network) are publicly available from their respective repositories.

Conflicts of Interest: The author declares no conflict of interest.

References

1. [1] European Commission. Industry 5.0: Towards a Sustainable, Human-Centric and Resilient European Industry; Publications Office of the European Union: Luxembourg, 2021.
2. [2] Xu, X.; Lu, Y.; Vogel-Heuser, B.; Wang, L. Industry 4.0 and Industry 5.0—Inception, conception and perception. J. Manuf. Syst. 2021, 61, 530–535.
3. [3] Leng, J.; Sha, W.; Wang, B.; Zheng, P.; Zhuang, C.; Liu, Q.; Wuest, T.; Mourtzis, D.; Wang, L. Industry 5.0: Prospect and retrospect. J. Manuf. Syst. 2022, 65, 279–295.
4. [4] Javaid, M.; Haleem, A.; Singh, R.P.; Suman, R. Industry 5.0: Potential applications in COVID-19. J. Ind. Integr. Manag. 2020, 5, 507–530.
5. [5] Maddikunta, P.K.R.; Pham, Q.-V.; Prabadevi, B.; Deepa, N.; Dev, K.; Gadekallu, T.R.; Ruby, R.; Liyanage, M. Industry 5.0: A survey on enabling technologies and potential applications. J. Ind. Inf. Integr. 2022, 26, 100257.
6. [6] Cloud Security Alliance. Identity Is the Cornerstone of Zero Trust; CSA Research Report; Cloud Security Alliance: Seattle, WA, USA, 2025.
7. [7] Rose, S.; Borchert, O.; Mitchell, S.; Connelly, S. Zero Trust Architecture; NIST SP 800-207; National Institute of Standards and Technology: Gaithersburg, MD, USA, 2020.
8. [8] Thales Group. Zero Trust for Critical Infrastructure Security; Thales White Paper; Thales: Paris, France, 2024.

 doi:10.20944/preprints202603.1618.v1

9. [9] ISACA. Adaptive Identity Is the Future of IAM and Zero Trust Alone Won't Get Us There; ISACA Now Blog, 2025.
10. [10] Industrial Cyber. Industrial IAM Emerges as Next Battleground in Cyber Defense Amid Legacy and Operational Hurdles; Industrial Cyber: Washington, DC, USA, 2025.
11. [11] Yang, A.; Li, M.; Zhang, Y. A Novel Zero-Trust Identity Framework for Agentic AI: Decentralized Authentication and Fine-Grained Access Control. arXiv 2025, arXiv:2505.19301.
12. [12] European Union. Regulation (EU) 2016/679 (General Data Protection Regulation); Official Journal of the European Union: Brussels, Belgium, 2016.
13. [13] Torra, V. Data Privacy: Foundations, New Developments and the Big Data Challenge; Springer: Cham, Switzerland, 2017.
14. [14] Fortinet. 2025 State of Operational Technology and Cybersecurity Report; Fortinet: Sunnyvale, CA, USA, 2025.
15. [15] Industrial Cyber. 2026 and Beyond: Urgent Need for Integrated Cybersecurity Strategies in Evolving Industrial Landscape; Industrial Cyber: Washington, DC, USA, 2025.
16. [16] CISA. Zero Trust Maturity Model, Version 2.0; Cybersecurity and Infrastructure Security Agency: Washington, DC, USA, 2023.
17. [17] Department of Defense. DoD Zero Trust Strategy and Roadmap; U.S. Department of Defense: Washington, DC, USA, 2022.
18. [18] LeCun, Y. A Path Towards Autonomous Machine Intelligence. OpenReview 2022, Version 0.9.2.
19. [19] Peffers, K.; Tuunanen, T.; Rothenberger, M.A.; Chatterjee, S. A Design Science Research Methodology for Information Systems Research. J. Manag. Inf. Syst. 2007, 24, 45–77.
20. [20] Chandramouli, R. A Zero Trust Architecture Model for Access Control in Cloud-Native Applications in Multi-Cloud Environments; NIST SP 800-207A; National Institute of Standards and Technology: Gaithersburg, MD, USA, 2023.
21. [21] Thales Group. 2025 Data Threat Report: Critical Infrastructure Edition; Thales: Paris, France, 2025.
22. [22] Moustafa, N.; Keshk, M.; Turnbull, B. Cybersecurity in Industry 5.0: Open Challenges and Future Directions. arXiv 2024, arXiv:2410.09538.
23. [23] Ahmad, R.; Alsmadi, I.; Alhamdani, W. Securing Industry 5.0: An Explainable Deep Learning Model for Intrusion Detection in Cyber-Physical Systems. Comput. Electr. Eng. 2025, 123, 110104.
24. [24] Bello, S.A.; Oyedele, L.O.; Akinade, O.O. Factors Impacting Cybersecurity Transformation: An Industry 5.0 Perspective. J. Bus. Res. 2024, 180, 114723.
25. [25] Rajawat, A.S.; Goyal, S.B.; Chauhan, R.; Verma, C. Fusion of AI and Blockchain for Securing Industry 5.0 Supply Chains. IEEE Access 2024, 12, 45823–45841.
26. [26] Assran, M.; Duval, Q.; Misra, I.; Bojanowski, P.; Vincent, P.; Rabbat, M.; LeCun, Y.; Ballas, N. Self-Supervised Learning from Images with a Joint-Embedding Predictive Architecture. In Proceedings of the IEEE/CVF CVPR, Vancouver, BC, Canada, 2023; pp. 15619–15629.
27. [27] Bardes, A.; Ponce, J.; LeCun, Y. V-JEPA: Latent Video Prediction for Visual Representation Learning. arXiv 2024, arXiv:2404.16930.
28. [28] Groth, J. On the Size of Pairing-Based Non-interactive Arguments. In Advances in Cryptology—EUROCRYPT 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 305–326.
29. [29] Gabizon, A.; Williamson, Z.J.; Ciobotaru, O. PLONK: Permutations over Lagrange-bases for Oecumenical Noninteractive Arguments of Knowledge. IACR Cryptol. ePrint Arch. 2019, 2019/953.
30. [30] Kothapalli, A.; Setty, S.; Tzialla, I. Nova: Recursive Zero-Knowledge Arguments from Folding Schemes. In Advances in Cryptology—CRYPTO 2022; Springer: Cham, Switzerland, 2022; pp. 247–277.
31. [31] Li, J.; Wang, X.; Zhang, Y. A Blockchain-Based Privacy-Preserving Data Governance Framework for Industry 5.0 Smart Factories. Peer-to-Peer Netw. Appl. 2025, 18, 42.
32. [32] Dwork, C.; Roth, A. The Algorithmic Foundations of Differential Privacy. Found. Trends Theor. Comput. Sci. 2014, 9, 211–407.
33. [33] McMahan, H.B.; Moore, E.; Ramage, D.; Hampson, S.; y Arcas, B.A. Communication-Efficient Learning of Deep Networks from Decentralized Data. In Proceedings of the AISTATS 2017; PMLR: Fort Lauderdale, FL, USA, 2017; pp. 1273–1282.

 doi:10.20944/preprints202603.1618.v1

34. [34] Xiao, X.; Tao, Y. Output perturbation with query relaxation. Proc. VLDB Endow. 2008, 1, 857–869.
35. [35] Yampolskiy, R.V.; Govindaraju, V. Behavioural biometrics: A survey and classification. Int. J. Biom. 2008, 1, 81–113.
36. [36] Alsultan, A.; Warwick, K.; Wei, H. Non-conventional keystroke dynamics for user authentication. Pattern Recognit. Lett. 2017, 89, 53–59.
37. [37] Antal, M.; Nemes, G. The MOBIKEY Keystroke Dynamics Password Database: Benchmark Results. Softw. Pract. Exp. 2016, 46, 999–1023.
38. [38] Delgado-Santos, P.; Tolosana, R.; Guest, R.; Vera-Rodriguez, R.; Deravi, F.; Morales, A. A Survey of Privacy Vulnerabilities of Mobile Device Sensors. ACM Comput. Surv. 2022, 54, 1–30.
39. [39] Dolev, D.; Yao, A. On the security of public key protocols. IEEE Trans. Inf. Theory 1983, 29, 198–208.
40. [40] MITRE. ATT&CK for ICS; MITRE Corporation: McLean, VA, USA, 2024.
41. [41] Google Cloud. Cybersecurity Forecast 2026; Google: Mountain View, CA, USA, 2025.
42. [42] Greshake, K.; Abdelnabi, S.; Mishra, S.; Endres, C.; Holz, T.; Fritz, M. Not What You've Signed Up For: Compromising Real-World LLM-Integrated Applications with Indirect Prompt Injection. In Proceedings of the AISec 2023; ACM: Copenhagen, Denmark, 2023; pp. 79–90.
43. [43] Tishby, N.; Pereira, F.C.; Bialek, W. The Information Bottleneck Method. In Proceedings of the 37th Allerton Conference; University of Illinois: Monticello, IL, USA, 1999; pp. 368–377.
44. [44] Wald, A. Sequential Analysis; John Wiley & Sons: New York, NY, USA, 1947.
45. [45] Desfontaines, D.; Pejó, B. SoK: Differential Privacy in Natural Language Processing. In Proceedings of the PoPETs 2023; IACR: Lausanne, Switzerland, 2023; pp. 229–248.
46. [46] SimPy Development Team. SimPy: Discrete-Event Simulation for Python, Version 4.1. Available online: https://simpy.readthedocs.io (accessed on 15 January 2026).
47. [47] Glasser, J.; Lindauer, B. Bridging the Gap: A Pragmatic Approach to Generating Insider Threat Data. In Proceedings of the IEEE Security & Privacy Workshops; IEEE: San Jose, CA, USA, 2013; pp. 98–104.
48. [48] Turcotte, M.J.M.; Kent, A.D.; Hash, C. Unified Host and Network Data Set. In Data Science for Cyber-Security; World Scientific: Singapore, 2018; pp. 1–22.
49. [49] National Security Agency. CNSA Suite 2.0 and Quantum Computing FAQ; NSA: Fort Meade, MD, USA, 2023.