

AI-Driven Dynamic Trust in Zero Trust Architecture: A Systematic Literature Review of Machine Learning, Federated Learning and Governance Challenges

Melissa Sharon Tubili¹, Belinda Ndlovu²

^{1,2} Department of Informatics and Analytics, National University of Science and Technology, Bulawayo, Zimbabwe

Received:

June 4, 2026

Revised:

July 21, 2026

Accepted:

September 10, 2026

Published:

October 24, 2026

Corresponding Author:

Name*:

Belinda Ndlovu

Email*:

belinda.ndlovu@nust.ac.zw

DOI:

10.63158/journalisi.v8i5.1815

© 2026 Journal of Information Systems and Informatics. This open access article is distributed under a (CC-BY License)



Abstract. Advanced cyber threats have exposed the limitations of perimeter-based security models, increasing interest in Zero Trust Architecture (ZTA). This systematic literature review examines the integration of Artificial Intelligence (AI) and Machine Learning (ML) in ZTA, including dynamic trust scoring, behavioral analytics, and federated learning. Following PRISMA 2020, 22 studies published between 2022 and 2025 were identified from IEEE Xplore, ACM Digital Library, SpringerLink, MDPI, and Nature/Scientific Reports. The studies were appraised using a five-criterion quality instrument and synthesized narratively. Ensemble and hybrid approaches predominated, appearing in 14 of 22 studies. Although 12 studies reported quantitative results, only eight used real named datasets. Reported accuracies of 95%–99% were obtained under different datasets, tasks, threat models, and evaluation settings; therefore, they are not directly comparable or representative of a field-level benchmark. Only three studies compared adaptive learned trust with static rule-based baselines, and no study reported live operational deployment. Thirteen technique categories, eleven implementation barriers, and seven emerging trends were identified. This study proposes an integrated eight-component framework for AI-driven dynamic trust management and a five-domain research agenda. The conceptual framework requires empirical validation before operational adoption.

Keywords: Zero Trust Architecture, artificial intelligence, machine learning, dynamic trust scoring, federated learning.

1. INTRODUCTION

Advanced persistent threats, insider threats, ransomware, and supply chain attacks have demonstrated the limitations of conventional network security solutions based on the perimeter model [1]. Traditional security models assume traffic within the network perimeter is trusted, and today's sophisticated threats exploit this assumption too often through lateral movement, privilege escalation, and credential theft. Cloud services, mobile users, and IoT technologies have made the notion of a network perimeter obsolete [2] [3].

Zero Trust Architecture (ZTA) responds by never trusting and always verifying: every access request is authenticated, authorised, and continuously validated regardless of network location [2], [3], [4], [5]. By continuously evaluating identity, device posture, application context, and network behaviour, ZTA moves security controls from the static perimeter to every individual request, making AI-driven trust computation the natural enforcement mechanism.

Because the term is used inconsistently, this paper defines dynamic trust at the outset. It denotes a continuously recomputed, machine-derived confidence value attached to a subject (user, device, workload, or session) that a policy decision point uses to grant, restrict, step up, or revoke access during a session, rather than only at initial authentication. Four functions are commonly bundled under the term and are treated as distinct throughout: real-time trust scoring, which produces the value; behavioural risk calculation, which supplies its features; policy adaptation, which changes the rule set the score is evaluated against; and continuous authorisation, which acts on the score during the session. This review covers all four and states in each results section which is at issue. Operationally, the score enters enforcement in one of three ways observed across the included studies: as a threshold gating access, as a continuous input to risk-adaptive policy varying authentication level or permission scope, or as a trigger for session termination and re-authentication when the score falls mid-session.

Several adjacent concepts are frequently conflated and are distinguished here. Zero Trust Architecture is the security model of per-request verification and least privilege; Zero Trust Network Access is an access-brokering technology implementing part of it; dynamic

access control is the enforcement mechanism; intrusion detection identifies malicious activity; and behavioural authentication verifies identity from behavioural signals. The key distinction here is output and consumer: an intrusion detection system raises an alert for a human analyst, whereas AI-ZTA dynamic trust produces a score that a policy decision point consumes automatically to change what a subject may do. Reviews of AI-driven intrusion detection algorithms and their applications [6], [7] illustrate the former category, in which the output is an alert for human adjudication rather than an access decision. A study is therefore included only where the AI or ML component informs an access or trust decision within a Zero Trust framework, not where it performs detection or malware classification in isolation.

The principle is sound but difficult to operate. Static, rule-based ZTA implementations struggle with the scale, velocity and variability of modern access patterns, cannot adapt to behaviour that changes within a session, and generate administrative overhead that grows with policy complexity [8], [9]. This gap between AI and ML is what this work aims to fill [10], [11].

Existing reviews have examined Zero Trust Architecture [2], [12], AI-driven intrusion detection [13], [14], [15], federated learning [16], [17], and behavioural authentication [18] individually. Table 1 compares this review with the closest prior syntheses along the dimensions that determine coverage and shows that none of the reviews compared their addresses AI-powered dynamic trust mechanisms, behavioural analytics, federated learning, implementation barriers and performance outcomes within a single quality-stratified synthesis. This is stated as a coverage gap identified against the reviews listed in Table 1, rather than as a claim of priority over the literature.

Table 1. Coverage of this review compared with the closest prior reviews

Prior review	Scope	Not covered
[2]	Zero Trust principles and architecture	AI/ML focus; quantitative synthesis; barriers
[12]	Zero Trust adoption challenges	AI treated in passing; no quality stratification
[13], [14]	AI-driven intrusion detection	Trust scoring; policy adaptation; ZTA scoping

[17]	Federated learning for cybersecurity	Not scoped to Zero Trust
[18]	Behavioural authentication	All other dimensions
This review	All five dimensions together, stratified by evidence type and study quality	Exhaustive algorithmic taxonomy

The review is global in scope. Its findings apply to any sector adopting Zero Trust, and they are of relevance to regulated and resource-constrained environments, including financial services in developing economies, where the barriers identified in Section 3.5, namely dataset scarcity, integration cost and skills availability, bind most tightly [19], [20]. No country-specific analysis is undertaken, and none is claimed. The research questions of this as follows:

1. What are the state-of-the-art approaches to integrating AI and ML technologies within a Zero Trust architecture?
2. How do AI-powered dynamic trust systems compare with traditional rule-based approaches regarding their effectiveness for security and operational efficiency?
3. What are the main barriers and constraints to the implementation of intelligent Zero Trust?
4. What are the emerging trends and future research directions related to AI-powered dynamic trust in Zero Trust Architecture specifically?

Each research question is answered in a dedicated results section: RQ1 in Section 3.2, RQ2 in Section 3.3, RQ3 in Section 3.5 and RQ4 in Section 3.6, and each is revisited in the discussion in Section 3. Together, these dimensions shape the governance and deployment challenges the discussion addresses.

2. METHODOLOGY

The review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to ensure the procedure was transparent and reproducible [21]. The PRISMA guidelines were selected because they outline how empirical research is included in a review and are widely used in computing and information systems reviews [22] [23].

2.1. Database Search Strategy

This process was carried out using five online databases: IEEE Xplore, ACM Digital Library, SpringerLink, MDPI, and Nature/Scientific Reports. These databases were selected for their broad indexing in cybersecurity, data science, and computer networking. The following Boolean expression was used across all databases:

("Zero Trust" OR "Zero Trust Architecture" OR "ZTA") AND ("Artificial Intelligence" OR "Machine Learning" OR "Deep Learning" OR "Federated Learning") AND ("Dynamic Trust" OR "Behavioral Analytics" OR "Anomaly Detection" OR "Intrusion Detection" OR "Cybersecurity")

Searches were executed between 1 and 30 December 2025, so the study reflects literature indexed up to 30 December 2025. Records published later in 2025 are not represented. The string above is the canonical form. Because the databases do not share a query syntax, we adapted it per platform while preserving the same three concept blocks and Boolean logic. In IEEE Xplore, it was entered in Command Search across All Metadata; in the ACM Digital Library, over Title, Abstract, and Author Keywords; in MDPI, over Title, Abstract, and Keywords. SpringerLink does not support Boolean nesting at this depth, so we decomposed the query into paired concept-block searches over title and abstract and combined the results manually, removing duplicates during deduplication. We used no truncation or wildcard operators, since support differs across platforms and their use would make retrieval non-comparable.

Nature/Scientific Reports is a publisher platform rather than a bibliographic database, and is reported as such. We searched it separately by title, abstract, and keywords because Scientific Reports publishes relevant applied cybersecurity work that is not consistently indexed elsewhere. Counting it as a fifth database in the PRISMA figures is a simplification, noted here for reproducibility. IEEE Xplore: 22 records. ACM Digital Library: 9 records. SpringerLink: 18 records. MDPI: 29 records. Nature/Scientific Reports: 12 records. Total: 90 records, of which 16 duplicates were removed to yield 74 unique records for screening. This breakdown is also shown in the PRISMA flow diagram (Figure 1).

2.2. Inclusion and Exclusion Criteria

Table 2 presents the eligibility criteria. To be included, a study had to examine the role of AI or ML in ZTAs, be published between 2022 and 2025, and be peer-reviewed. By limiting our analysis to 2022-2025, we focus on recent research while maintaining methodological sophistication. Two included studies, [24] and [25], could not be verified as peer-reviewed at the time of writing, as noted in their reference entries. Both are retained for transparency and excluded from the sensitivity synthesis in Section 3.4, so neither contributes to any quantitative finding.

Table 2. Inclusion and exclusion criteria

Criterion	Inclusion	Exclusion
Time Frame	2022-2025 (recent AI-ZTA advances)	Studies published before 2022
Language	English only	Non-English publications
Document Type	Peer-reviewed journal articles and conference proceedings	Grey literature, editorials, book chapters, unpublished theses
Publication Status	Fully published and peer-reviewed	Preprints, abstracts, or draft papers
Research Area	AI/ML integration in ZTA; behavioral analytics; federated learning for cybersecurity threat detection	Studies outside AI-ZTA scope or addressing unrelated security domains
Keywords	Zero Trust, Machine Learning, AI, dynamic trust, federated learning, behavioral authentication, anomaly detection, intrusion detection	Studies not explicitly addressing AI-powered trust evaluation or ZTA frameworks

2.3. Eligibility and Screening

The search returned 90 records. After removing 16 duplicates, 74 records remained for title and abstract screening; 38 were excluded for not addressing AI or ML within a Zero Trust framework, not meeting the language criterion, or falling outside the window. Of the 36 sought for full-text assessment, 14 were excluded for lacking an AI-ZTA trust or access decision, for being conceptual without analysis, or for duplicate reporting, leaving 22 studies for synthesis. Figure 1 shows the flow.

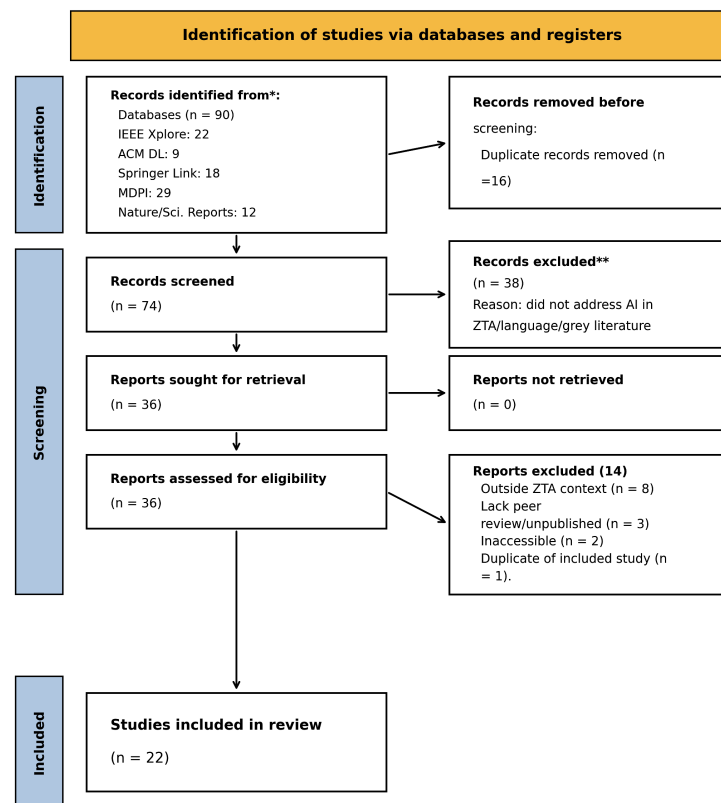


Figure 1. PRISMA 2020 flow diagram of the study selection process

2.4. Data Extraction Framework

Each included study was evaluated against a standardised extraction instrument of eight fields derived from the four research questions: country and year, level of evidence, AI/ML methodology (RQ1), quantitative metrics (RQ2), barriers (RQ3), trends (RQ4), opportunities, and the ZTA context of use. Quantitative metrics were recorded only where the study generated them; borrowed or illustrative figures are flagged in Table 4.

2.5. Quality Assessment of Included Studies

To assess the quality and strength of evidence from the studies included in the review, we assigned a score based on five elements: clarity of aims, methodology, data sources, rigour of evaluation, and limitations. Each element had a score of 1 if it was fulfilled and 0 if it was not, up to a maximum of 5. Each criterion was scored 1 if satisfied, and 0 if not. Objectives: a specific research aim is stated. Methodology: the approach is described in enough detail to be understood, including architecture where applicable. Data sources: the dataset is named, or for conceptual work the basis of analysis is stated. Evaluation rigour: quantitative results with stated metrics and conditions, or systematic argued

analysis where none are performed. Limitations: constraints on the study's own findings are identified.

Both authors conducted the appraisal independently and compared criteria across all 22 studies. Disagreements concerned Evaluation rigour for conceptual studies and Data sources for studies naming a dataset without version or partition and were resolved against a written rule applied consistently thereafter: conceptual studies score 1 on Evaluation rigour only where the analysis is systematic and traceable to stated premises, and a dataset must be named and publicly identifiable to score 1. Agreement before reconciliation was 91 per cent (100 of 110 judgements); the remainder were resolved by consensus. A formal Cohen's κ coefficient was not computed; the 91 per cent pre-reconciliation agreement (100 of 110 criteria-by-study judgements) is reported as a reliability approximation.

Studies [25] and [26] were retained rather than excluded, since excluding on quality grounds after screening would change the eligibility criteria post hoc, and both contribute to the thinly represented reinforcement learning category. Section 3.4 reports a sensitivity analysis excluding them. Risk of bias and confidence. Evaluation-setting bias arises because four of the twelve empirically evaluated studies report simulation only; reporting bias because near-ceiling results on saturated benchmarks are more publishable, inflating apparent field performance; and metric bias because figures are borrowed from cited sources rather than generated, which we flag in Table 4 and exclude from the sensitivity synthesis. Confidence is moderate for the technique taxonomy and barrier synthesis, and low for any comparative performance claim. Studies scoring 4 or 5 were treated as high quality, 3 as moderate, and 2 or below as low. Table 3 records 15 high, 5 moderate and 2 low.

Table 3. Quality assessment of the included studies

Study	Objectives	Method	Dataset	Evaluation	Limitations	Total	Quality
[27]	1	1	1	1	1	5	High
[28]	1	1	0	1	1	4	High
[29]	1	1	1	1	1	5	High
[30]	1	1	1	1	1	5	High
[31]	1	1	0	1	1	4	High

Study	Objectives	Method	Dataset	Evaluation	Limitations	Total	Quality
[24]	1	1	0	1	1	4	High
[32]	1	1	0	0	1	3	Moderate
[33]	1	1	0	0	1	3	Moderate
[34]	1	1	0	0	1	3	Moderate
[25]	1	0	0	0	0	1	Low
[26]	1	0	0	0	0	1	Low
[35]	1	1	0	1	1	4	High
[36]	1	1	1	1	1	5	High
[37]	1	1	1	1	1	5	High
[38]	1	1	1	1	1	5	High
[39]	1	1	0	1	1	4	High
[40]	1	1	0	0	1	3	Moderate
[41]	1	1	0	1	1	4	High
[42]	1	1	0	0	1	3	Moderate
[43]	1	1	1	1	1	5	High
[44]	1	1	0	1	1	4	High
[45]	1	1	1	1	1	5	High

Metrics reported in [31] are inconsistent in the abstract and tables in that paper; therefore, all quantitative claims in that paper are provisional. Papers [25] and [26] have low-quality ratings because they use secondary metrics and do not conduct any evaluations.

3. RESULTS AND DISCUSSION

3.1. Characteristics of the Included Studies

The 22 included studies [24] to [45] were published between 2022 and 2025, most in 2024 and 2025, and span consumer IoT, industrial control, cloud and enterprise, healthcare and 5G/6G contexts. Eight report original empirical evaluations on real named datasets, four

evaluate in simulation, five are conceptual or architectural, and the remainder combine design with case-study validation. Table 4 presents the full data extraction for all 22 studies, mapping each to its evidence type, AI/ML approach, reported metrics, and ZTA deployment context.

Table 4. Main synthesis: study, authors/country/year, evidence type, AI/ML approach, reported metrics, and ZTA context across all 22 included studies

Stud y	Authors / Country / Year	Evidence Type	AI/ML Approach	Reported Metrics	ZTA Context & Key Barriers
[27]	Prajwalasimha et al. India 2025	Empirical (real)	Transformer + GNN + LLM + Federated- learning fuzzy ensemble	Acc 96.4%; FPR 2.3%; AUC 0.93; Dataset: Proprietary	General ZTA. Barriers: complexity; integration cost
[28]	Bradatsch et al. Germany 2023	Empirical (sim.)	Subjective-Logic trust scoring vs additive baseline; dynamic threshold	Score- distribution & computation- time eval; Dataset: Synthetic (enterprise)	Enterprise network ZTA. Barriers: threshold standardisatio n
[29]	Munasinghe et al. Sri Lanka 2023	Empirical (real)	Nine ML classifiers + stacking + online learning	Acc >97%; FPR <2%; F1 0.97; online adaptation; Dataset: UNSW- NB15	Network ZTA. Barriers: concept drift; retraining
[30]	Cao et al. China 2025	Empirical (real)	PCA + AHP + dynamic behaviour-aware framework, RF/SVM	Acc 96.76%; FPR 1.9%; F1 0.96; Latency <20 ms; Dataset: CICIDS2017	Mobile device access control. Barriers: feature engineering; latency
[31]	Shrivastava et al. India 2025	Empirical (sim.)	AI-ZTA, LSTM-RNN + blockchain	Acc 95%; dwell 80 to 20 min; 25 TPS;	Energy/SCADA. Barriers: data

Study	Authors / Country / Year	Evidence Type	AI/ML Approach	Reported Metrics	ZTA Context & Key Barriers
				Dataset: Synthetic (ICS)	realism; synthetic-only
[24]	Meyers et al. n.r. 2025	Empirical (sim.)	Behavioral biometrics + RNN/autoencoder/GMM	Acc >97%; Latency <150 ms; mimicry resist. 92%; Dataset: Synthetic (biometric)	Continuous auth. Barriers: dataset realism; latency
[32]	Rehan USA 2022	Conceptual	Multi-cloud ZTA design review	None	Multi-cloud. Barriers: interoperability
[33]	Dang & Liu China 2025	Conceptual	Blockchain + AI + digital-twin ZTA	None	Smart campus. Barriers: deployment scale; interoperability
[34]	Dalal USA 2025	Conceptual	Narrative ZTA for distributed networks	None	Distributed networks
[25]	Selvaprasanth India 2025	Borrowed metrics	Weighted risk score + Q-learning	Borrowed, unverified	Cloud-native. Barriers: validation difficulty
[26]	Mangla n.r. 2024	Borrowed metrics	RF/SVM/CNN/XGBoost/ RL to policy mapping	Sample Data (illustrative)	Policy enforcement. Barriers: validation difficulty
[35]	Laghari et al. China/Pakistan 2025	Empirical (sim.)	AI/ML/DL + federated learning zero-trust IDS	Simulation-based evaluation vs state-of-the-	Industrial IoT (IIoT) ZTA. Barriers: dataset specification;

Study	Authors / Country / Year	Evidence Type	AI/ML Approach	Reported Metrics	ZTA Context & Key Barriers
				art; Dataset: not named	real-time deployment
[36]	Nagarajan et al. India 2024	Empirical (real)	Deep CNN-BiLSTM + Bayes trust score (AIDL-XTS)	Acc 98.2%; F1 0.97; Latency <25 ms; EER & auth time measured; Dataset: WISDM-HARB; HMOG; UCI-HAR	Consumer IoT; user/device/app auth. Barriers: heterogeneous -device scalability
[37]	Nawshin et al. Qatar 2024	Empirical (real)	DP-RFECV-FNN (differential privacy + FNN + feature selection)	Acc 96.84%; F1 0.96; privacy loss 0.5; 89.5% at 1.0; Dataset: Android malware	IoT ZTA (healthcare wearables). Barriers: privacy-utility trade-off; budget
[38]	Javeed et al. Pakistan/Australia 2024	Empirical (real)	Federated CNN-BiLSTM (FL-ZTA IDS)	Acc >=98%; FPR <2%; F1 0.98; 20 edge devices; 10 FL rounds; Dataset: CICIDS2017; Edge-IIoTset	IoT ZTA; FL privacy. Barriers: concept drift; edge heterogeneity
[39]	Xie et al. China 2024	Empirical (sim.)	Dynamic federated learning + blockchain (IIoT)	Convergence & availability gains; Dataset: Synthetic (IIoT)	Industrial IIoT ZTA. Barriers: on-chain aggregation; scalability
[40]	Hussain et al. Australia 2024	Conceptual + case study	Federated learning + ZTA architectural integration	Case-study validation; Dataset: Case study (no benchmark)	Cloud/BYOD enterprise. Barriers: interoperability

Study	Authors / Country / Year	Evidence Type	AI/ML Approach	Reported Metrics	ZTA Context & Key Barriers
					y; remote work
[41]	Pokhrel et al. Australia 2024	Empirical (sim.)	Blockchain-based FL + unsupervised anomaly detection	Acc degrades >50% malicious nodes; convergence vs BFL; Dataset: Synthetic (IoT)	Decentralised IoT ZTA. Barriers: poisoning resilience; scalability
[42]	Ramezanzpour & Jagannath USA 2022	Conceptual	i-ZTA: DNN, Bayesian, Fuzzy, GNN for 5G/6G O-RAN	Latency/accuracy trade-offs (theoretical)	5G/6G O-RAN ZTA. Barriers: real-time latency; network slicing
[43]	Jeong et al. South Korea 2025	Empirical (real)	Trust-Score model: 4-Factor weighted sum	Acc 95.9%; F1 0.97; Latency <20 ms; 20 sub-metrics; sensitivity analysis; Dataset: UNSW-NB15	Enterprise ZTA access control. Barriers: threshold determination; non-linear effects
[44]	Hassan et al. Pakistan/Saudi Arabia 2025	Empirical (sim.)	Isolation-Forest UEBA + ZTA SIEM/SOAR; adaptive playbooks	Latency MTTR <10 s (network); 15-25 s (insider); SOC automation validated; Dataset: Synthetic (enterprise)	Enterprise ZTA; SOC automation. Barriers: static-rule SIEM gap
[45]	Xu et al. China 2024	Empirical (real)	Lightweight secure FL (residual self-attention CNN + pruning + encryption)	Acc ~99%; >=89.6% comms-cost reduction;	Industrial wireless ZTA (ICS). Barriers: edge FL;

Study	Authors / Country / Year	Evidence Type	AI/ML Approach	Reported Metrics	ZTA Context & Key Barriers
				Dataset: NSL-KDD; UNSW-NB15; Smart Meter	communications overhead

3.2. AI and ML Approaches for Dynamic Trust (RQ1)

Thirteen distinct categories of AI or ML approaches are identified across the twenty-two studies. Federated learning is the most common supporting technology, in seven studies [27], [35], [38], [39], [40], [41], [45], followed by deep learning hybrids in four [36], [38], [43], [45] and blockchain integrity layers in four [31], [33], [39], [41]. [29] applies nine ML classifiers with stacking and online learning, and [42] is a conceptual i-ZTA analysis, so neither is federated. Seven is used consistently in Table 7, Section 3.6 and Section 3.13.

Ensemble and hybrid designs dominate in fourteen studies (64%). The most elaborate is the transformer-GNN-LLM-federated ensemble of [27]. The simplest models are the 4-factor weighted trust score of [43] and the Subjective-Logic approach of [28], both of which emphasise computational efficiency and interpretability. This range, from highly complex to highly simple, suggests that the trade-off between accuracy and latency is being considered at design time, although none of the included studies quantifies it.

Three conceptual or architectural studies [32], [34], [42] describe AI-driven trust at the design level without training a model or evaluating data. Study [42] is the most thorough, mapping AI engine types to specific ZTA functions for 5G/6G O-RAN environments; this is useful for architectural planning but is not empirical validation.

Table 5. AI and ML techniques reported, grouped by category (22 studies)

Category	Techniques reported	Studies
Feature Engineering / Preprocessing	Principal Component Analysis (PCA); Analytic Hierarchy Process (AHP); Dynamic behaviour-aware Feature Framework; Recursive Feature Elimination with CV (RFECV); Mutual Information Feature selection	[30], [37], [38]

Category	Techniques reported	Studies
Classical / Supervised ML	Random Forest (RF); Support Vector Machine (SVM); Multi-classifier stacking ensemble (9 classifiers); XGBoost; Isolation Forest (UEBA anomaly detection)	[29], [30], [26], [38], [44], [45]
Deep and Sequence Models	Transformer; Graph Neural Network (GNN); Large Language Model (LLM) in trust engine; LSTM-RNN; Recurrent Neural Network (RNN); Autoencoder; Gaussian Mixture Model (GMM); Convolutional Neural Network (CNN); Bidirectional LSTM (BiLSTM); Residual self-attention CNN + model pruning; Deep Neural Network (DNN); Feedforward Neural Network (FNN); Bayesian classifier	[27], [31], [24], [26], [36], [37], [38], [42], [45]
Reinforcement Learning	Q-learning / RL trust adaptation	[25], [26]
Online / Adaptive Learning	Online (incremental) learning; Unsupervised anomaly detection (clustering)	[29], [41]
Fuzzy, Weighted and Formal Trust Scoring	Fuzzy trust inference engine; Weighted risk score; Bayes-theorem trust score (Bayes-TSC); 4-factor weighted trust score (TS model); Subjective-Logic trust scoring	[27], [28], [25], [36], [43]
Federated and Distributed Learning	Federated Learning (FL), general; Dynamic Federated Learning + blockchain aggregation; Federated CNN-BiLSTM (FL-IDS); Lightweight secure FL (pruning + encryption); Blockchain-based robust FL aggregation	[27], [35], [38], [39], [40], [41], [45]
Privacy-Preserving Mechanisms	Differential Privacy (DP); Encryption within FL communication	[37], [45]
Distributed Integrity and Architectural Layers	Blockchain; Digital Twin; SIEM / SOAR integration; i-ZTA O-RAN architecture (5G/6G); Conceptual AI-ZTA design only	[31], [32], [33], [34], [39], [40], [41], [42], [44]

3.3. Quantitative Performance and Comparison with Rule-Based Control (RQ2)

Twelve of the twenty-two studies evaluate their mechanism quantitatively, eight on real named datasets and four in simulation. Table 4 records each report's metrics. Of the eight

studies evaluated on real data, seven use publicly available named benchmarks (UNSW-NB15, CICIDS2017, NSL-KDD, Edge-IIoTset, and comparable corpora) and are, in principle, reproducible. Study [27] uses a proprietary dataset that is not named, described, or publicly available; its reported 96.4 per cent is not independently verifiable, not comparable with public benchmark results, and is excluded from the sensitivity synthesis. Across the eight empirically evaluated studies, accuracy ranges from 95.9 per cent [43] to about 99 per cent [45], with F1 between 0.96 and 0.98 and reported latencies under 25 ms where measured. These are individual results under individual conditions and are not aggregated.

Three studies directly compare AI-based and static policies in terms of control effectiveness. Research [29], [30], and [43] state that the AI-based adaptive system shows higher efficiency than static policies, while [29] and [30] also report lower false-positive rates. Even more important for the entire research set is the ability to adapt behaviour. Static approaches assume stable user behaviour and threat prediction. In contrast, the AI-based approach continuously updates its trust model based on behaviour, context, and environmental changes.

Notes: [Real] = evaluation on a real, named, publicly available, or disclosed dataset. [Sim.] = evaluation on synthetic or simulation data. -- = metric not reported by the study.

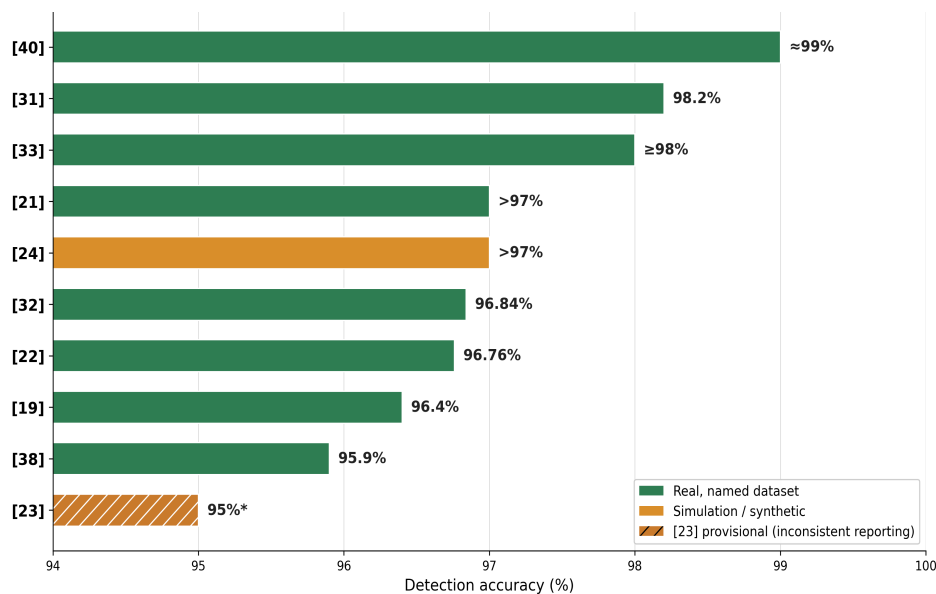


Figure 2. Detection accuracy of empirically evaluated studies, shaded by evaluation setting

The shading distinguishes evaluation settings, and the categories are not equivalent forms of evidence. Bars representing simulation-based results and the bar representing the proprietary dataset of [27] should not be read as commensurate with results obtained on real, publicly available benchmarks. Do not compare bar heights across datasets and tasks. The figure summarises what each study reports; it is not a ranking.

3.4. Sensitivity Synthesis

Because the included studies mixed evidence of markedly different strength, we recomputed the principal quantitative findings after removing the weakest evidence. We applied three exclusions: the two low-quality studies [25] and [26]; the four studies evaluated only in simulation [35], [39], [41] and [44]; and study [27], whose proprietary dataset is not verifiable. This leaves six studies evaluated on real, publicly available datasets with self-generated metrics: [29], [36], [37], [38], [43] and [45].

Within this restricted set, accuracy ranges from about 96.8 to 99 per cent, narrower than the range across all included studies but still spanning four datasets and different tasks, and therefore still not directly comparable. Ensemble and hybrid designs remain predominant. The claim that adaptive AI trust outperforms static rule-based control is materially weakened, since only two of the four direct comparisons survive the exclusions. The barrier synthesis is largely unaffected, as it is descriptive across all evidence types.

The number of studies supporting each principal conclusion after exclusions is therefore: technique predominance of ensemble and hybrid designs, six of six; superiority of learned adaptive trust over static rule-based control, two studies; near-ceiling accuracy on public benchmarks, six studies; live operational deployment, zero studies. These figures, rather than the counts across all included studies, should be used when weighing the strength of any claim made in this review.

3.5. Implementation Barriers (RQ3)

Eleven implementation barriers emerged, ranked by frequency in Table 6. Dataset scarcity and reliance on synthetic data lead, reported in nine studies [31],[25],[26],[24],[35],[39],[41],[42],[44], followed by integration complexity with legacy infrastructure and the accuracy-latency tension. Four barrier categories appear here that

earlier, narrower reviews did not capture: the privacy-utility trade-off in federated and differentially private systems, reported by four studies; adversarial and poisoning attacks on FL aggregation, newly documented by [39]; threshold determination and standardization for trust scoring, raised by [43] and [28]; and communications overhead in edge-deployed FL systems, tackled directly by [45]. As the field matures technically, its problems move from model accuracy towards system-level concerns such as deployment, governance, and adversarial robustness.

Table 6. Implementation barriers reported across the 22 included studies, ranked by frequency

Barrier	Studies Reporting	Count
Dataset scarcity / reliance on synthetic or borrowed data	[31], [25], [26], [24], [35], [39], [41], [42], [44]	9
Integration and computational complexity	[27], [30], [36], [42], [45]	5
Interoperability across heterogeneous environments	[32], [33], [38], [40], [42]	5
Privacy-utility trade-off in federated / privacy-preserving systems	[37], [38], [40], [41]	4
Latency budget for real-time trust scoring	[30], [24], [42], [43]	4
Validation and benchmarking difficulty	[25], [26], [41], [44]	4
Scalability with increasing device/node count	[33], [38], [39], [41]	4
Concept drift and need for continuous retraining	[29], [38], [40]	3
Deployment scale in resource-constrained or OT environments	[33], [31], [35], [39]	4
Threshold determination and lack of standardisation	[43], [28]	2
Poisoning / adversarial attacks on FL aggregation	[41]	1

3.6. Emerging Trends (RQ4)

Seven emerging trends emerged, ranked in Table 7. Federated and distributed learning dominate in seven studies, having moved from an emerging technique to a near-

mainstream enabler of privacy-preserving ZTA. Blockchain for auditable trust decisions appears in five studies, indicating growing attention to accountability. Two trends are not addressed by the reviews compared in Table 1. Privacy-preserving mechanisms, specifically differential privacy and encrypted federated communications, now form a category of their own and address the privacy-utility trade-off more precisely than federated learning alone. Formal, interpretable trust scoring is the second, offering mathematically grounded models that could support regulatory accountability.

Table 7. Emerging trends across the 22 included studies, ranked by frequency

Emerging Trend	Supporting Studies	Count	Research Questions Addressed
Federated and distributed learning	[27], [35], [38], [39], [40], [41], [45]	7	RQ1, RQ4
Blockchain for auditable / tamper-evident trust	[27], [33], [31], [39], [41]	5	RQ4
Continuous and behavioral authentication	[30], [24], [36], [43]	4	RQ1, RQ2
Privacy-preserving mechanisms (DP, encryption)	[37], [38], [45]	3	RQ1, RQ4
Reinforcement learning for adaptive trust policy	[25], [26]	2	RQ1, RQ4
Digital twin and simulation-driven design	[33], [31]	2	RQ4
LLM / Generative AI in ZTA trust engines	[27]	1	RQ1, RQ4

3.7. Opportunities Identified

Opportunities reported across the included studies cluster on three themes. The most frequent, in nine studies, is adaptive AI trust outperforming static rule-based control, subject to the evidential qualification in Section 3.4. Federated and privacy-preserving learning for cross-organisational threat intelligence without data sharing follows in seven studies, and lightweight or interpretable trust scoring for latency-constrained

enforcement in four. These opportunities mirror the barriers in Table 6: each names a constraint the included studies expect AI to relieve rather than a demonstrated result. This claim is confined to the three studies that compare directly within the same experimental setting ([29],[30],[43]); the remaining six studies in the nine-count identify this as an expected benefit rather than a demonstrated result. The reviews do not address two opportunity categories compared in Table 1: explainable, interpretable trust decisions (XAI), which [43] and [28] flag as a key enabler of regulatory compliance and stakeholder trust; and communications-efficient FL for resource-constrained IoT, which [38] and [45] show is achievable.

This section interprets the findings of Section 3, compares studies, and assesses the strength of the evidence supporting each claim, organised by the four research questions. Three constraints govern the comparisons below. Reported accuracies are not pooled or ranked, because studies use different datasets, tasks, threat models and definitions of trust. Any statement that adaptive AI trust outperforms static rule-based control is confined to the four studies making that comparison within a single experimental setting. Several high accuracies come from UNSW-NB15, CICIDS2017, and NSL-KDD, near-ceiling for years, so upper-nineties values may reflect benchmark saturation rather than operational capability. No study reports live deployment, so no claim here rests on operational evidence.

The reviewed studies comprise [24] to [45]. All other references are contextual literature and are not part of that set. They situate the findings and support the research agenda. They were either not retrieved by the search string or did not meet the Table 2 criteria, most commonly because they do not integrate AI or ML with a Zero Trust access decision, and they appear in no PRISMA count, table or figure. All citations with numbers [1]–[23] and [46]–[60] in this section are contextual literature outside the SLR corpus and appear in no PRISMA count, table or figure.

Regarding AI and ML approaches to dynamic trust (RQ1), the trend across the included studies is towards ensemble and hybrid designs rather than single models, the most complex being the transformer-GNN-LLM-federated ensemble of [27] and the simplest the weighted trust score of [43] and the Subjective-Logic approach of [28], which prioritise efficiency and interpretability. Two approaches remain under-explored: formal

trust scoring [28], [43], which addresses computational complexity with inspectable models rather than more layers, and differential privacy as a design tier, applied directly in training by [37] with the resulting privacy-utility cost reported.

Maturity is uneven once approaches are judged by how rigorously they were realised. The most architecturally ambitious proposals [27], [33], [42] are not the best validated: [33] remains conceptual and the ensemble in [27] is evaluated on a proprietary dataset. Architectural ambition and evidential strength are not correlated in these studies, a pattern also visible in contemporaneous work outside it on ML-driven threat detection in ZTA [46], [48], [7], SIEM-integrated pipelines [49], ML-based compliance evaluation [50] and game-theoretic trust [51].

Turning to quantitative performance and comparison with rule-based control (RQ2), within the constraints stated above, AI-driven dynamic trust is directionally supported across the included studies and empirically supported in a minority. Among the eight real-data studies, reported accuracy falls between 95.9 and 99 percent with F1 between 0.96 and 0.98, and [45] reaches approximately 99 percent while reporting a communications-cost reduction of at least 89.6 percent, indicating that accuracy and efficiency are not necessarily opposed. These figures describe individual studies under their own conditions and are not aggregated.

The evidence base carries deficiencies that limit these figures. Studies [25] and [26] lack independent assessment, [31] presents metrics borrowed from cited sources rather than generated by their own evaluation, and four studies report only simulation results. Section 3.4 reports the synthesis excluding these. The eleven implementation barriers identified (RQ3) are best understood as symptoms of a structurally constrained field rather than isolated technical problems. Each barrier listed in Table 6 has been encountered before. Together, all eleven reveal a state of the art constrained more by research practices and surrounding infrastructure than by any technological limit.

The accuracy-latency tension is the most structurally important technical barrier. The highest-accuracy studies [27], [38] run the heaviest architectures, while the lowest-latency designs [28], [43] accept lower accuracy for inline enforcement. None of the

studies explicitly quantify this trade-off, which is a specific and tractable gap for future work.

Privacy-utility trade-off in federated and differentially private systems [37], [38], [41], [42] is a newer barrier. Nawshin et al. [37] report accuracy falling from 96.84 to 89.5 per cent as the privacy budget tightens, quantifying a tension between privacy compliance and security efficacy that prior ZTA reviews have not documented. Collectively, the barriers are structural rather than technical, and the same pattern appears in Zero Trust maturity frameworks [52] and sector deployments in enterprise analytics [53], power distribution [54] and electric transport [55]. The biggest obstacle is the lack of standardised evaluation.

The seven emerging trends identified (RQ4) collectively describe a research community deepening its architectural ambition. All seven trends in Table 7 represent increasing architectural ambition. Federated learning (seven studies), blockchain (five studies), continuous behavioural authentication (four studies) and differential privacy (three studies) are common enough to qualify as near-mainstream elements of AI-ZTA design. Reinforcement learning for adaptive policy and digital twins, with two studies each, remain emergent.

The most significant recent development is the adoption of privacy-preserving mechanisms beyond federated learning, notably differential privacy [37] and encryption-based model compression [45], [47], which reposition privacy from an incidental property of federated architecture to a primary design requirement with formal guarantees. Countering the trend toward complexity, studies [43] and [28] trade some accuracy for interpretability, lower computational cost, and sensitivity analysis, which may matter more for deployment than marginal accuracy gains.

A clear gap across the included studies is the limited attention to explainability. Several studies achieve high trust classification accuracy, but few explain how a trust score is computed or why an access decision was reached, which matters increasingly given regulatory interest in algorithmic accountability [56] and the emphasis placed on transparency and documented risk management in national AI governance frameworks [57], [20]. Compounding complexity is the associated risk: each added layer, whether

blockchain, digital twin or LLM, increases latency, attack surface and the difficulty of explaining a denial, and none of the included studies evaluates that cumulative cost.

RQ1 (AI/ML approaches: thirteen technique sub-types in nine grouped categories; ensemble and hybrid designs predominate (14 of 22 studies); federated learning is the most common supporting technology (7 studies). RQ2 (performance vs rule-based control): eight studies report real-dataset accuracy of 95.9–99%, with F1 0.96–0.98; only three directly compare against a static baseline; no study establishes operational superiority. RQ3 (barriers): nine studies report dataset scarcity; eleven barriers identified; accuracy-latency trade-off and absence of evaluation standards are the most structurally important. RQ4 (trends): federated learning (7 studies) and blockchain (5) are near-mainstream; privacy-preserving mechanisms beyond FL and formal interpretable trust scoring are trends not addressed by prior reviews in Table 1.

Across all four research questions, three cross-cutting synthesis gaps emerge. Three gaps recur across the four research questions. The first is the absence of an agreed evaluation standard: studies report different metrics on different datasets, so no cumulative performance picture emerges. The second is the absence of operational validation: no study reports deployment in a live Zero Trust environment, so behaviour under real traffic, real users and real adversaries is unknown. The third is weak governance and explainability, which are named as requirements far more often than they are implemented. The literature shows AI-powered Zero Trust moving from an access-control mechanism to a broader adaptive security capability, with trust continuously evaluated through behavioural analytics, federated intelligence, differential privacy, and autonomous decision support. The performance results are promising and increasingly well-grounded, but evaluation practices remain inconsistent, and real-world deployments remain rare. The next phase of research needs to prioritise rigorous validation, explainability, governance, and benchmark standardization over further architectural invention.

An integrated framework for AI-driven dynamic trust management can be derived inductively from the recurring patterns across [24] to [45]. Across the twenty-two studies, five architectural patterns recur regardless of technique or deployment domain: continuous multi-source behavioural signal acquisition; AI-enabled trust inference;

dynamic trust score generation; policy-driven enforcement; and a feedback loop for adaptation. From these, the paper proposes an eight-component integrated framework for AI-driven dynamic trust management, derived inductively from recurring themes across [24] to [45] and offered both as a synthesis of what these studies contain and as a specification of what they lack.

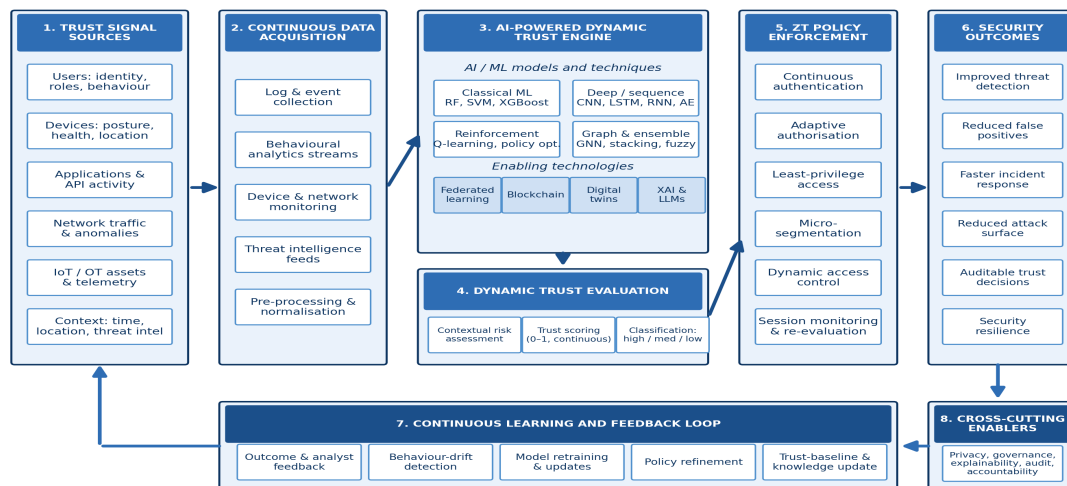


Figure 3. Integrated framework for AI-driven dynamic trust management in Zero Trust environments, inductively derived from the synthesis of the included studies

Table 8 sets out each component and the validation required before operational use.

Table 8. Framework components and validation requirements

Component	Inputs to outputs	Validation required
1 Behavioural signal sources	User interaction telemetry, device posture, application access patterns, network flows to raw event streams	Coverage audit against assets in scope; signal availability in target environment
2 Acquisition and normalisation	Raw event streams to normalised, time-aligned feature vectors	Ingestion latency and loss rate under production load
3 AI trust inference engine	Feature vectors to continuous trust value per subject and session	Accuracy, FP and FN rates on operational data; inference latency on target hardware

Component	Inputs to outputs	Validation required
4 Trust score generation	Model outputs, trust history, contextual risk to calibrated, bounded score	Calibration and stability; no oscillation under normal behaviour
5 Policy decision point	Trust score, resource sensitivity, policy rules to grant, restrict, step up or deny	Decision consistency; false-denial rate on legitimate activity
6 Enforcement and continuous authorisation	Policy decisions to enforced session state, mid-session revocation	Enforcement latency; revocation within a stated bound
7 Feedback and adaptation	Enforcement outcomes, analyst adjudications, drift signals to retraining triggers, policy updates	Drift detection sensitivity; no degradation over time
8 Governance and accountability	All decisions, model versions, data flows to audit trails, explanations, privacy accounting, appeal records	Auditability against applicable regime; decision reconstructable and appealable

The framework synthesises recurring patterns across the included studies and the design requirements implied by its gaps. It has not been implemented or evaluated, and Table 8 states what such an evaluation would need to measure. Component 8, governance and accountability, underlies the rest and is the least represented in the included studies, so it is treated as the principal gap.

Because Component 8 is the principal gap, it is operationalised rather than named. Governance requires five capabilities: decision traceability, so any score producing a denial can be reconstructed with its features, model version and threshold; explanation at enforcement, stating which behavioural factors moved the score at a level an analyst can use; human oversight and appeal, with a defined override route that is logged and fed back into evaluation; model lifecycle governance covering versioning, drift monitoring and retraining approval, since a silently degrading model is an access-control failure; and privacy accounting, recording what data is collected, on what lawful basis,

for how long, and where differential privacy applies, the budget consumed. Frameworks for responsive and responsible AI set out the same requirements from the governance side, treating privacy, consent and accountability as design obligations rather than downstream compliance activities [58], [59]. These capabilities are absent or partial in the included studies. No included study reports an appeal mechanism for automated access denial and only [37] reports an explicit privacy budget. Component 8 is therefore proposed based on what the included studies lack rather than what they demonstrate, and this is stated as a limitation of the framework.

The contributions of this review carry direct practical implications: For architects, ensemble and hybrid designs dominate the literature but carry a latency cost no study reconciles with production requirements; lightweight formal trust-scoring approaches [43], [28] are the more realistic starting point, where enforcement is inline and published accuracy should be read as a laboratory ceiling. For security operations teams, alert volume, analyst workload and false-denial rates under production traffic are unmeasured in this literature, so a phased rollout in monitor-only mode is warranted. For regulated sectors, the governance capabilities above bind before detection performance does: finance requires a traceable and appealable denial path, healthcare, where ensemble methods have been applied to threat and incident detection [60], [15], requires conservative thresholds and human override, and in critical infrastructure the integration barriers of Section 3.5 bind earlier than any model-performance limit.

The framework contributes in three ways: it draws together research streams usually treated separately, including machine learning, behavioural analytics, federated learning, differential privacy, reinforcement learning, formal trust scoring and governance; it makes the governance layer explicit as a component rather than an afterthought; and it states, in Table 8, what each component would have to demonstrate before operational use. It synthesises existing knowledge and provides a starting point for empirical validation, not a validated design.

The synthesis identifies five priority domains for future work; each tied to a gap in the cross-study analysis. No references are cited here, since these are directions rather than claims about existing literature. Public benchmark datasets are needed that are built for continuous trust scoring in Zero Trust environments rather than repurposed from

intrusion detection. Existing corpora predate the problem and encourage benchmark saturation rather than progress. Explainable and transparent trust. Trust decisions must be explainable at the point of enforcement, in terms a security analyst can act on and a subject can contest, and explanations should be evaluated against their effect on human decisions rather than assumed useful. Federated, privacy-preserving and lightweight intelligence. Open problems include scalability under non-independent and non-identically distributed data, resilience to model poisoning, communication overhead on constrained devices, and the governance of collaboratively trained models. Autonomous and agentic trust management. Where systems observe risk and adapt policy with minimal human intervention, the research questions concern the bounds of delegated authority, the conditions for human intervention, and how autonomous adaptation is audited. Governance, ethics and regulatory compliance. As access decisions are delegated to learned models, work is needed on accountability structures, fairness in access outcomes, and the evidence a regulated organisation must retain to justify an automated denial. Across all five, the priority is operational field evidence. A single well-documented deployment reporting false-denial rates, analyst workload and enforcement latency would contribute more than further benchmark comparison.

4. CONCLUSION

This systematic review synthesised twenty-two studies published between 2022 and 2025 on integrating Artificial Intelligence and Machine Learning into dynamic trust mechanisms within Zero Trust Architecture. The included studies converge on ensemble and hybrid designs (fourteen of twenty-two), thirteen AI or ML technique categories led by federated learning (seven studies), eleven implementation barriers, and seven emerging trends; yet persistent weaknesses remain: two studies report only borrowed or illustrative metrics, five are purely conceptual, four evaluate only in simulation, and none report deployment in a live operational environment. The review makes three contributions: a quality-stratified, PRISMA-guided synthesis of the AI-ZTA evidence base covering dynamic trust mechanisms, behavioural analytics, federated learning, barriers and performance outcomes in a single quality-stratified analysis not addressed by the prior reviews compared in Table 1; an inductively derived eight-component integrated framework set out in Figure 3 and Table 8; and a five-domain research agenda grounded in the cross-cutting gaps identified in the synthesis. These contributions are bounded by

six limitations: English-only searching; five sources, one of which (Nature/Scientific Reports) is a publisher platform rather than a database; a corpus of twenty-two studies, which is small for field-level claims; heterogeneous metrics that prevent meta-analysis; retention of two low-quality studies addressed by the sensitivity analysis in Section 3.4; and the absence of live deployment evidence. The 95 to 99 per cent accuracy range describes individual studies under their own conditions and is neither a field-level benchmark nor evidence of production readiness: only three studies compare directly against a static rule-based baseline, and several high accuracies come from benchmarks that have been near-saturated for years. The proposed framework is correspondingly conceptual and synthesis-derived; it has not been implemented, evaluated, or validated and should not be treated as an operational specification until empirically tested. Future work should prioritise standardised benchmarks built for continuous trust scoring rather than repurposed intrusion-detection corpora; operational field trials reporting false-denial rates, analyst workload, and enforcement latency alongside accuracy; adversarial evaluation including model poisoning in federated settings and evasion of behavioural trust scoring; and explainability of trust decisions at the point of enforcement. As access-control decisions are increasingly delegated to learned models, governance, explainability, and human oversight cease to be desirable additions and become preconditions for legitimate deployment: a system that can deny a clinician access to a patient record, or a customer access to their funds, must be able to explain that decision, log it, and allow a human to review and reverse it.

REFERENCES

- [1] B. D. Lund, T. H. Lee, Z. Wang, T. Wang, and N. R. Mannuru, "Zero Trust Cybersecurity: Procedures and Considerations in Context," *Encyclopedia*, vol. 4, no. 4, pp. 1520–1533, Dec. 2024, doi: 10.3390/encyclopedia4040099.
- [2] N. F. Syed, S. W. Shah, A. Shaghaghi, A. Anwar, Z. Baig, and R. Doss, "Zero Trust Architecture (ZTA): A Comprehensive Survey," *IEEE Access*, vol. 10, pp. 57143–57179, 2022, doi: 10.1109/ACCESS.2022.3174679.
- [3] C. Liu *et al.*, "Dissecting Zero Trust: Research Landscape and Its Implementation in IoT," *Cybersecurity*, vol. 7, no. 1, p. 20, 2024, doi: 10.1186/s42400-024-00212-0.

- [4] A. Dube, B. Mpande, F. Dzehonye, and T. Chazuza, "Zero Trust Architecture: Redefining Security," in *Lecture Notes in Networks and Systems*, Springer Nature Switzerland, 2026, pp. 26-39, doi: 10.1007/978-3-032-20533-9.
- [5] S. Rose, O. Borchert, S. Mitchell, and S. Connelly, "Zero Trust Architecture," National Institute of Standards and Technology, Gaithersburg, MD, NIST SP 800-207, Aug. 2020. doi: 10.6028/NIST.SP.800-207
- [6] M. Moyo, B. Ndlovu, M. Moyo, and T. Siwela, "AI-Driven Intrusion Detection Systems: Algorithms, Key Applications, Efficacy, Opportunities and Ethical Implications," in *Advances in Micro-Electronics, Embedded Systems and IoT. Lecture Notes in Electrical Engineering*, vol. 1600. Springer, Cham, 2026, doi: 10.1007/978-3-032-20533-9_4
- [7] K. L. Chizengwe and B. Ndlovu, "A Systematic Review of Agentic AI for Threat Detection and Mitigation in 5G Networks," *J. Inf. Syst. Informatics*, vol. 8, pp. 315–354, 2026, doi: 10.63158/journalisi.v8i1.1382.
- [8] F. A. Qazi, "Study of Zero Trust Architecture for Applications and Network Security," in *Proc. IEEE 19th Int. Conf. Smart Communities: Improving Quality of Life Using ICT, IoT and AI (HONET)*, 2022. doi: 10.1109/HONET56683.2022.10019186.
- [9] N. Kashyap, V. Kumar, and D. Baweja, "Zero Trust Architecture: Evaluating its Impact on Modern Information Security," *Institute of Electrical and Electronics Engineers (IEEE)*, Jan. 2026, pp. 1–6. doi: 10.1109/iccica67008.2025.11337489.
- [10] D. Ajish, "The significance of artificial intelligence in zero trust technologies: a comprehensive review," *J. Electr. Syst. Inf. Technol.*, vol. 11, no. 1, 2024, doi: 10.1186/s43067-024-00155-z.
- [11] E. S. Hosney, I. T. A. Halim, and A. H. Yousef, "An Artificial Intelligence Approach for Deploying Zero Trust Architecture (ZTA)," in *Proc. 5th Int. Conf. Computing and Informatics (ICCI)*, 2022. doi: 10.1109/ICCI54321.2022.9756117.
- [12] A. I. Weinberg, "Zero Trust Implementation in the Emerging Technologies Era: A Survey," *Complex Eng. Syst.*, vol. 4, p. 16, 2024, doi: 10.20517/ces.2024.41.
- [13] M. Rahmati and A. Pagano, "Federated Learning-Driven Cybersecurity Framework for IoT Networks with Privacy Preserving and Real-Time Threat Detection Capabilities," *Informatics*, vol. 12, no. 3, Sep. 2025, doi: 10.3390/informatics12030062.
- [14] W. W. Lo, S. Layeghy, M. Sarhan, and M. Portmann, "Online Self-Supervised Deep Learning for Intrusion Detection Systems," *IEEE Trans. Inf. Forensics Secur.*, 2024, doi: 10.1109/TIFS.2024.3402148.

- [15] B. Ndlovu, "AI-Driven Intrusion Detection in Healthcare Systems: A Systematic Review of Techniques, Evaluation Practices and Deployment Readiness, with a Reusable Assessment Framework," vol. 3, pp. 1–38, 2026, doi: 10.63158/IJAIS.v3i1.58.
- [16] J. L. Hernandez-Ramos *et al*, "Intrusion Detection Based on Federated Learning: A Systematic Review," Jul. 12, 2025, *Association For Computing Machinery*. doi: 10.1145/3731596.
- [17] A. Vyas, P.-C. Lin, R.-H. Hwang, and M. Tripathi, "Privacy-Preserving Federated Learning for Intrusion Detection in IoT Environments: A Survey," *IEEE Access*, vol. 12, pp. 127018–127050, 2024, doi: 10.1109/ACCESS.2024.3454211.
- [18] C.-J. Chew, P.-Y. Wang, and J.-S. Lee, "Behavioural Analysis Zero-Trust Architecture Relying on Adaptive Multifactor and Threat Determination," *KSII Trans. Internet Inf. Syst.*, vol. 17, no. 9, pp. 2529–2549, 2023, doi: 10.3837/tiis.2023.09.012.
- [19] V. Maphosa, "An overview of cybersecurity in Zimbabwe's financial services sector," *F1000Research*, vol. 12, p. 1251, 2024, doi: 10.12688/f1000research.132823.2.
- [20] C. M. Mukwaira, Y. Chigariro, and B. Ndlovu, "An Integrated Cybersecurity Governance Ecosystem for Healthcare: A Quality-Appraised Systematic Review of Governance, Risk, and Compliance Frameworks for AI, Cloud, and Connected Health Technologies," vol. 13, pp. 617–636, 2026, doi: 10.15294/sji.v13i3.49854.
- [21] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement," *PLoS Med*, vol. 6, no. 7, p. e1000097, 2009, doi: 10.1371/journal.pmed.1000097.
- [22] M. J. Page and others, "The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews," *BMJ*, vol. 372, p. n71, 2021, doi: 10.1136/bmj.n71.
- [23] K. Maguraushe and B. M. Ndlovu, "The Use of Smart Technologies for Enhancing Palliative Care: A Systematic Review," *Digit. Heal.*, vol. 10, p. 20552076241271836, 2024, doi: 10.1177/20552076241271835.
- [24] J. R. Meyers, A. D. Thompson, M. L. Greene, S. K. Roberts, and C. James, "AI-Based Continuous Authentication in Zero Trust Environments," Jun. 2025.
- [25] P. Selvaprasanth, "Artificial Intelligence-Driven Access Control for Resilient Security Frameworks in Cloud-Native Environments," 2025.
- [26] M. Mangla, "Integrating Machine Learning with Zero Trust Principles for Real-Time Threat Detection and Response," *Iconic Res. Eng. Journals*, vol. 7, no. 9, pp. 564–575, Mar. 2024.

- [27] S. N. Prajwalasimha, A. Pimpalkar, N. Shelke, and D. K. J. B. Saini, "Zero Trust Architectures Empowered by AI: A Paradigm Shift in Cloud and Edge Cybersecurity," in *Proceedings of 3rd International Conference on Sustainable Computing and Data Communication Systems, ICSCDS 2025*, Institute of Electrical and Electronics Engineers Inc., 2025, pp. 328–335. doi: 10.1109/ICSCDS65426.2025.11166875.
- [28] L. Bradatsch, O. Miroshkin, N. Trkulja, and F. Kargl, "Zero Trust Score-based Network-level Access Control in Enterprise Networks," in *2023 IEEE 22nd International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, 2023, pp. 1422–1429. doi: 10.1109/TrustCom60117.2023.00194.
- [29] S. Munasinghe, N. Piyarathna, E. Wijerathne, U. Jayasinghe, and S. Namal, "Machine Learning Based Zero Trust Architecture for Secure Networking," in *2023 IEEE 17th International Conference on Industrial and Information Systems, ICIIIS 2023 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 365–370. doi: 10.1109/ICIIIS58898.2023.10253610.
- [30] X. Cao, Z. Chen, M. Yang, X. Wang, and X. Wu, "Zero Trust-Based Dynamic and Continuous Access Control for Mobile Devices," in *Proc. IEEE 24th Int. Conf. Trust, Security and Privacy in Computing and Communications (TrustCom)*, 2025. doi: 10.1109/Trustcom66490.2025.00344.
- [31] A. Shrivastava *et al*, "Tackling Cyber Espionage in Critical Energy Sectors Using AI-Driven Zero Trust Systems," Institute of Electrical and Electronics Engineers (IEEE), Dec. 2025, pp. 355–359. doi: 10.1109/inspire67328.2025.11300601.
- [32] H. Rehan, "Zero-Trust Architecture for Securing Multi-Cloud Environments," *Cybersecurity Netw. Def. Res.*, vol. 2, no. 2, pp. 236–273, 2022.
- [33] D. Lina and L. Jiannan, "A Multi-Layered Cybersecurity Framework for Smart Campuses: Integrating Blockchain, AI, Zero Trust Architecture, and Digital Twins," in *2025 IEEE 3rd International Conference on Control, Electronics and Computer Technology, ICCECT 2025*, Institute of Electrical and Electronics Engineers Inc., 2025, pp. 11–14. doi: 10.1109/ICCECT64621.2025.11339781.
- [34] A. Dalal, "Designing Zero Trust Security Models to Protect Distributed Networks and Minimize Cyber Risks," *Int. J. Manag. Technol. Eng.*, vol. 11, no. 11, pp. 142–164, 2025.
- [35] A. A. Laghari *et al*, "A novel and secure artificial intelligence enabled zero trust intrusion detection in industrial internet of things architecture," *Sci. Rep.*, vol. 15, no. 1, pp. 1–15, 2025, doi: 10.1038/s41598-025-11738-9.

- [36] M. Nagarajan, G. G. Devarajan, M. S. Thangakrishnan, T. V. Ramana, A. K. Bashir, and A. A. AlZubi, "Artificial Intelligence Based Zero Trust Security Approach for Consumer Industry," *IEEE Trans. Consum. Electron.*, vol. PP, p. 1, Aug. 2024, doi: 10.1109/TCE.2024.3412772.
- [37] F. Nawshin, D. Unal, M. Hammoudeh, and P. Suganthan, "AI-powered malware detection with Differential Privacy for zero trust security in Internet of Things networks," *Ad Hoc Networks*, vol. 161, p. 103523, Apr. 2024, doi: 10.1016/j.adhoc.2024.103523.
- [38] D. Javeed, M. S. Saeed, M. Adil, P. Kumar, and A. Jolfaei, "A Federated learning-based zero trust intrusion detection system for Internet of Things," *Ad Hoc Networks*, vol. 162, p. 103540, 2024, doi: 10.1016/j.adhoc.2024.103540.
- [39] H. Xie, Y. Wang, Y. Ding, C. Yang, H. Liang, and B. Qin, "Industrial Wireless Internet Zero Trust Model: Zero Trust Meets Dynamic Federated Learning with Blockchain," *Wirel. Commun.*, vol. 31, no. 2, pp. 22–29, Apr. 2024, doi: 10.1109/MWC.001.2300368.
- [40] M. Hussain, S. Pal, Z. Jadidi, E. Foo, and S. Kanhere, "Federated Zero Trust Architecture using Artificial Intelligence," *IEEE Wirel. Commun.*, vol. 31, no. 2, pp. 30–35, 2024, doi: 10.1109/MWC.001.2300405.
- [41] S. R. Pokhrel, L. Yang, S. Rajasegarar, and G. Li, "Robust Zero Trust Architecture: Joint Blockchain based Federated learning and Anomaly Detection based Framework," in *Proceedings of the SIGCOMM Workshop on Zero Trust Architecture for Next Generation Communications*, in ZTA-NextGen '24. New York, NY, USA: Association for Computing Machinery, 2024, pp. 7–12. doi: 10.1145/3672200.3673878.
- [42] K. Ramezanpour and J. Jagannath, "Intelligent zero trust architecture for 5G/6G networks: Principles, challenges, and the role of machine learning in the context of O-RAN," *Comput. Networks*, vol. 217, p. 109358, 2022, doi: 10.1016/j.comnet.2022.109358.
- [43] E. Jeong and D. Yang, "A Trust Score-Based Access Control Model for Zero Trust Architecture: Design, Sensitivity Analysis, and Real-World Performance Evaluation," 2025. doi: 10.3390/app15179551.
- [44] A. Hassan, A. Rauf, N. Shafqat, R. Latif, and H. Khan, "ZenGuard: A Machine Learning-Based Zero Trust Framework for Context-Aware Threat Mitigation Using SIEM, SOAR, and UEBA," *Sci. Rep.*, vol. 15, p. 35871, 2025, doi: 10.1038/s41598-025-20998-4.

- [45] C. Xu, X. Du, L. Li, X. Li, and H. Yu, "End-Edge Collaborative Lightweight Secure Federated Learning for Anomaly Detection of Wireless Industrial Control Systems," *IEEE Open J. Ind. Electron. Soc.*, vol. PP, pp. 1–11, Jan. 2024, doi: 10.1109/OJIES.2024.3370496.
- [46] H. Sedjelmaci, K. Tourki, and N. Ansari, "Enabling 6G Security: The Synergy of Zero Trust Architecture and Artificial Intelligence," *IEEE Netw.*, vol. PP, p. 1, Jan. 2023, doi: 10.1109/MNET.2023.3326003.
- [47] F. M. Dandure and B. Ndlovu, "Federated Learning for Privacy-Preserving Sentiment Analysis in Distributed Electronic Health Record Environments: A Systematic Literature Review," *J. Inf. Syst. Informatics*, vol. 8, pp. 1812–1842, 2026, doi: 10.63158/journalisi.v8i2.1546.
- [48] V. M. Zanpure, "Improving Zero Trust Architecture with Machine Learning: Automatic Threat Detection Using Network Traffic Analysis," in *Proc. 2nd Int. Conf. Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE)*, 2025. doi: 10.1109/RMKMATE64874.2025.11042837.
- [49] I. Stankov, "Comparing Azure Sentinel and ML-Extended Solutions Applied to a Zero Trust Architecture," in *Proc. 32nd National Conf. with International Participation (TELECOM)*, 2024. doi: 10.1109/TELECOM63374.2024.10812246.
- [50] H. Teymourlouei, "A Machine Learning Approach to the Evaluation of Zero Trust Compliance in Network Infrastructure," in *International Conference on Electrical, Computer, Communications and Mechatronics Engineering, ICECCME 2023*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ICECCME57830.2023.10253205.
- [51] A. A. Hossain and D. Ahner, "Game Theory–Driven AI for Zero Trust Security," in *2025 2nd International Conference on Artificial Intelligence, Metaverse, and Cybersecurity (ICAMAC)*, IEEE, Oct. 2025, pp. 1–6. doi: 10.1109/ICAMAC67779.2025.11398547.
- [52] S. Alnoaimi and A. Alomary, "Zero Trust Security: A Comprehensive Comparative Analysis of Zero Trust Maturity Models," in *2nd International Conference on IT Innovations and Knowledge Discovery, ITIKD 2024*, Institute of Electrical and Electronics Engineers Inc., 2025. doi: 10.1109/ITIKD63574.2025.11005097.

- [53] D. Narasappa, "Integrating Zero Trust Architecture with Automation and Analytics for Resilient Cybersecurity," in *Proc. 3rd Int. Conf. Data Science and Network Security (ICDSNS)*, 2025. doi: 10.1109/ICDSNS65743.2025.11168551.
- [54] S. Bai, W. Sheng, Y. Shang, G. Wang, L. Zhou, and L. Sun, "Research on Information Security Protection System of Distribution Network Based on Zero Trust Architecture," in *Proc. IEEE Int. Conf. Energy Internet (ICEI)*, 2024. doi: 10.1109/ICEI63732.2024.10917171.
- [55] G. Muriithi *et al.*, "Zero Trust Architecture for Electric Transportation Systems: A Systematic Survey and Deep Learning Framework for Replay Attack Detection," *IEEE Open J. Veh. Technol.*, vol. 6, pp. 2171–2194, 2025, doi: 10.1109/OJVT.2025.3592041.
- [56] A. Barredo Arrieta and others, "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, 2020, doi: 10.1016/j.inffus.2019.12.012.
- [57] National Institute of Standards and Technology, "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," Jan. 2023. doi: 10.6028/NIST.AI.100-1.
- [58] B. Ndlovu and K. Maguraushe, "Balancing Ethics and Privacy in the Use of Artificial Intelligence in Institutions of Higher Learning: A Framework for Responsive AI Systems," *IJIE (Indonesian J. Informatics Educ.)*, vol. 9, no. 1, p. 39, 2025, doi: 10.20961/ijie.v9i1.100723.
- [59] T. Kapuya, W. Kubiku, M. Dube, and T. Mukudu, "Ethical AI Frameworks: Balancing Privacy, Consent and Responsible Use," 7th European Conference on Industrial Engineering and Operations Management, Augsburg, Germany, July 16-18, 2024, doi: 10.46254/EU07.20240168.
- [60] A. Maramba and B. Ndlovu, "AI-Driven Threat and Incident Detection in Healthcare Cybersecurity: A Stacked Ensemble Model," in 2025 4th Zimbabwe Conf. Information and Communication Technology, 2025, pp. 1-9, doi: 10.1109/ZCICT67553.2025.11602039.