Journal of *Cybersecurity and Privacy*

MDPI

Review

The Erosion of Cybersecurity Zero-Trust Principles Through Generative AI: A Survey on the Challenges and Future Directions

Dan Xu [1,*], Iqbal Gondal [1], Xun Yi [1], Teo Susnjak [2], Paul Watters [3] and Timothy R. McIntosh [1,4]

[1] School of Computing Technologies, RMIT University, Melbourne, VIC 3000, Australia; iqbal.gondal@rmit.edu.au (I.G.); xun.yi@rmit.edu.au (X.Y.)
[2] School of Mathematical and Computational Sciences, Massey University, Auckland 0632, New Zealand
[3] Cyberstronomy Pty Ltd., Ballarat, VIC 3350, Australia
[4] Cyberoo Pty Ltd., Surry Hills, NSW 2010, Australia
* Correspondence: s3152414@student.rmit.edu.au

Abstract

Generative artificial intelligence (AI) and persistent empirical gaps are reshaping the cyber threat landscape faster than Zero-Trust Architecture (ZTA) research can respond. We reviewed 10 recent ZTA surveys and 136 primary studies (2022–2024) and found that 98% provided only partial or no real-world validation, leaving several core controls largely untested. Our critique, therefore, proceeds on two axes: first, mainstream ZTA research is empirically under-powered and operationally unproven; second, generative-AI attacks exploit these very weaknesses, accelerating policy bypass and detection failure. To expose this compounding risk, we contribute the *Cyber Fraud Kill Chain* (CFKC), a seven-stage attacker model (target identification, preparation, engagement, deception, execution, monetization, and cover-up) that maps specific generative techniques to NIST SP 800-207 components they erode. The CFKC highlights how synthetic identities, context manipulation and adversarial telemetry drive up false-negative rates, extend dwell time, and sidestep audit trails, thereby undermining the Zero-Trust principles of *verify explicitly* and *assume breach*. Existing guidance offers no systematic countermeasures for AI-scaled attacks, and that compliance regimes struggle to audit content that AI can mutate on demand. Finally, we outline research directions for adaptive, evidence-driven ZTA, and we argue that incremental extensions of current ZTA that are insufficient; only a generative-AI-aware redesign will sustain defensive parity in the coming threat cycle.

Keywords: zero trust; generative AI; cybersecurity; adversarial attacks; trust mechanisms; AI auditing

Received: 18 August 2025
Revised: 29 September 2025
Accepted: 9 October 2025
Published: 15 October 2025

Citation: Xu, D.; Gondal, I.; Yi, X.; Susnjak, T.; Watters, P.; McIntosh, T.R. The Erosion of Cybersecurity Zero-Trust Principles Through Generative AI: A Survey on the Challenges and Future Directions. *J. Cybersecur. Priv.* **2025**, *5*, 87. https://doi.org/10.3390/jcp5040087

1. Introduction

Zero trust is a foundational cybersecurity principle that advocates for the continuous verification of all entities, both inside and outside a network, and the rejection of implicit trust [1]. This principle challenges traditional perimeter-based security models, which have proven insufficient against modern cyber threats, particularly cyber fraud [2]. Cyber fraud and zero trust are closely interlinked; while cyber fraud exploits gaps in trust and access control, zero trust emerged as a necessary countermeasure [3]. The operationalization of zero trust has been formalized through frameworks like NIST SP 800-207 [4], which provides structured guidelines for implementation. Such frameworks are further realized in architectures such as the UK's National Cyber Security Centre (NCSC) Zero Trust Architecture Design Principles and Australia's "Essential 8". By focusing on verifying user

identity, contextual data, and device health before granting access to sensitive resources, zero-trust frameworks aim to minimize the risks of fraud and unauthorized access [2,5]. The principle of "never trust, always verify" offers a proactive and dynamic approach to mitigate the risks posed by an increasingly hostile cyber environment, making zero trust an essential cybersecurity strategy worthy of thorough investigation.

Despite its promise, zero trust has faced significant challenges in addressing the complexities of evolving cyber threats [2]. While it has been refined to include granular access controls, continuous monitoring, and real-time threat detection [1], it struggles with scalability, legacy system integration, and adapting to dynamic IT environments, such as cloud services and mobile devices [1,2,5]. These inherent issues have been compounded by the rise of generative AI, which enables attackers to create synthetic identities, manipulate contextual data, and automate sophisticated attacks [6]. Generative AI weaponizes three structural weak points in today's zero-trust deployments. First, identity governance pipelines still rely on human-verifiable documents and liveness checks, yet diffusion headshots and LLM-written résumés can now pass those gates unchallenged. Second, adaptive-trust algorithms are trained on clean, human-generated baselines, so a single poisoned burst of synthetic telemetry can drift their decision boundaries. Third, micro-segmented networks depend on deep-packet inspection and protocol heuristics, but GAN-shaped traffic can masquerade as sanctioned backups or VoIP streams. Because model updates arrive faster than security policies can be revised, rule sets and attestations age out in weeks, not years. Unless zero-trust architecture (ZTA) becomes generative-AI-aware, it risks sliding from a living security posture to a static checklist, widening the gap between policy intent and operational reality. The cybersecurity community must, therefore, critically assess and adapt ZTA models to meet emerging challenges. Throughout this article, we use the term "erosion of Zero Trust principles" to denote a quantifiable weakening (manifested as higher false-negative rates, longer attacker dwell time, or policy bypasses), rather than total collapse; the metrics and evidence supporting this definition are examined in Section 6.

This survey critically examines the intersection of zero trust and generative AI, focusing on how generative AI undermines the foundational principles of zero trust. Our study has been structured around five key research questions that explored the definition and implementation of trust within ZTA and the evolving impact of human, process, and technological factors. Through a two-step critical analysis, we have reviewed 10 recent surveys and 136 primary research studies. While some progress has been made, our analysis has revealed that many of these studies lacked sufficient empirical evidence or were not well executed, leaving significant research gaps. Such gaps were further exacerbated by generative AI, rendering many existing zero-trust defenses less effective or obsolete. Additionally, by introducing the novel "Cyber Fraud Kill Chain", we have provided a structured framework from the attacker's perspective to assess the full impact of generative AI on zero trust. Our framework has demonstrated how generative AI amplifies weaknesses in zero-trust models, making the gaps identified in current studies even more problematic, urgent, and complex. Our survey has highlighted the need for adaptive, generative-AI-aware zero-trust frameworks that address both existing and emerging vulnerabilities in modern cybersecurity environments. Throughout Sections 2–5, we flag how generative AI exacerbates existing ZTA gaps—in survey coverage (Section 2), architectural variations (Section 3), review methodology (Section 4), and empirical evaluations (Section 5)—culminating in our AI-aware "Cyber Fraud Kill Chain" (CFKC) in Section 6.

The three major contributions of this study are as follows:

1. We systematically evaluated zero-trust architecture (ZTA) research through the lens of the NIST SP 800-207 taxonomy, assessing 10 surveys and 136 primary studies published since 2022. This evaluation identified critical research gaps, including an

insufficient focus on behavior-based trust algorithms, continuous monitoring infrastructures, and enclave-based deployments, all of which are essential for addressing modern cybersecurity threats.

2. We demonstrated how generative AI amplifies vulnerabilities in existing ZTA models, such as eroding trust mechanisms, disrupting compliance processes, and automating sophisticated attack vectors. To address these challenges, we introduced the novel CFKC framework, which maps the stages of AI-driven fraud and provides actionable insights for improving ZTA defenses against such emerging threats.
3. We proposed a comprehensive set of research directions and practical recommendations to guide the evolution of zero-trust frameworks and architectures. These include integrating adaptive trust mechanisms, embedding AI-specific regulatory compliance, and prioritizing advanced defenses against generative AI-driven threats, thereby enabling ZTA to remain effective in modern cybersecurity environments.

2. Related Surveys

This section reviews existing surveys on ZTA to assess their contributions, methodologies, and limitations. The objective is to identify gaps in the current understanding of ZTA and justify the need for this survey, which builds on their limitations by providing a more recent and comprehensive evaluation of ZTA frameworks in the context of generative AI-driven threats. Our review of 10 prominent surveys [1,2,7–14] highlights diverse approaches to ZTA, with an emphasis on theoretical principles, technological advancements, and sector-specific implementations. However, these surveys collectively lack comprehensive empirical validation, real-world deployment examples, and practical guidance for addressing modern cybersecurity challenges. Theoretical Frameworks: Refs. [9,12] focus on the foundational principles and strategic importance of ZTA. While these studies provide valuable theoretical insights, they lack discussions on behavior-based trust mechanisms and continuous monitoring, which are critical for addressing advanced threats. Refs. [1,2] also explore strategic aspects of ZTA but offer limited coverage of real-world implementations or adaptive mechanisms for evolving attack vectors. Technological Advancements: Ref. [7] examines ZTA integration into 6G networks, highlighting continuous authentication and real-time risk assessment but neglecting practical deployment challenges. Ref. [10] provides a comparative analysis of ZTA within cloud computing, offering valuable insights into technology-specific issues. However, both studies lack empirical validation and detailed performance assessments. Practical Implementations: Refs. [11,14] explore ZTA in big data, IoT, and network environments, identifying key components and application areas. While these studies outline implementation strategies, they fail to address real-world challenges and evolving compliance requirements. Refs. [8,13] provide more recent evaluations but would benefit from case studies and empirical data to substantiate their findings. The summarized findings are presented in Table 1. Despite their breadth, none of these ten surveys consider the threat or defense implications of generative AI—an omission this paper remedies by systematically mapping AI-driven fraud vectors onto ZTA controls. Our analysis has highlighted a shared principal shortcoming: the predominant focus on theoretical principles over practical challenges, empirical validation, and real-world applications. We believe such gaps can arise from the difficulty of obtaining proprietary deployment data and the nascent stage of zero-trust implementations. Additionally, none of the reviewed surveys address the disruptive impact of generative AI on ZTA, creating a critical gap in understanding and mitigating emerging threats.

Table 1. Assessment of the 10 surveys on ZTA coverage (✓: comprehensive; Δ: partial or implicit) with structured synthesis additions.

Reference	Date	NIST Zero Trust Architecture 800-207				Consideration of AI Risk Management			Structured Synthesis			
		Variations of ZTA Approaches	Deployed Variations of the Abstract Architecture	Trust Algorithm	Network/Environment Components	ISO 42001	NIST AI RMF	EU AI Act	Domain/Scope	Methodology	Evidence/ Validation	Key Limitations
[11]	Q1 2020	✓		✓	✓				Big data, IoT, networks	Narrative survey	Limited (no deployments)	Minimal real-world challenges; no compliance evolution
[12]	Q4 2020	✓		Δ					Conceptual/theory	Conceptual review	None reported	Lacks behavioral trust and continuous monitoring
[9]	Q3 2021	✓		Δ	✓				Strategy/policy	Conceptual review	None reported	Strategic focus; no deployment guidance
[14]	Q2 2022	✓	Δ						IoT/big data	Narrative survey	Limited	Sparse performance analysis; integration gaps
[1]	Q2 2022	✓							Enterprise/general	Narrative survey	None reported	Little on adaptive controls or case studies
[10]	Q3 2022	✓	Δ		✓				Cloud computing	Comparative review	Limited	Missing empirical benchmarks and performance metrics
[7]	Q1 2023	✓	Δ	✓	✓				6G/mobile	Systematic-style survey	Limited	Deployment challenges underaddressed
[13]	Q4 2023	✓	✓	✓	✓				General/enterprise	Review	Partial (high-level)	Needs case studies and reproducible data
[8]	Q1 2024	✓		Δ	✓				Industry/ practitioner views	Multivocal review	Limited	Practitioner breadth; scarce empirical validation
[2]	Q2 2024	✓	Δ	✓					Governance/ verification	Position/verification-oriented survey	Limited	Narrow scope; limited deployment coverage
This survey		✓	✓	✓	✓	✓	✓	✓	Generative-AI risk to ZTA	Critical survey + mapping	Pilot evidence (Appendices)	Addresses AI-driven erosion; proposes CFKC and agenda

Note: The NIST SP 800-207 columns assess survey coverage of core ZTA principles. The final three columns evaluate whether the survey considers recent AI governance frameworks: ISO 42001 (AI management systems), NIST AI RMF (risk-based categorization and controls), and the EU AI Act (compliance and enforcement boundaries). These frameworks are not specific to ZTA, and nor are they tailored to generative-AI threats; they do not, by themselves, mitigate the erosion of ZTA assurances caused by adversarial AI. Our inclusion of these columns serves to highlight the absence of regulatory convergence, rather than to suggest that compliance with these instruments ensures ZTA effectiveness under AI-driven threat models. We, therefore, prioritize NIST SP 800-207 as the technical grounding for this survey while acknowledging the governance relevance of these newer instruments in Sections 2 and 7.

From such gaps, we derive the following five research questions (RQs) to guide our analysis. Through addressing them, this survey identifies critical gaps in existing ZTA research, demonstrates how generative AI amplifies vulnerabilities, and provides actionable insights for evolving ZTA frameworks to remain effective against modern threats.

1. **RQ1: How can "trust" and "zero trust" be properly defined within the context of ZTA?** A foundational understanding of trust mechanisms is essential for developing coherent ZTA frameworks. This RQ supports Contribution 1, which evaluates ZTA taxonomies and identifies gaps in trust algorithms.
2. **RQ2: What are the different ways to achieve effective implementation of ZTA across diverse operational environments?** Practical deployment strategies are critical for improving ZTA adaptability. This question aligns with Contribution 1 by addressing gaps in behavior-based trust and enclave deployments.
3. **RQ3: How have people factors, including user behavior, security culture, and insider threats, evolved in the context of ZTA implementation?** Generative AI introduces new complexities in user behavior and insider threat dynamics. This question aligns with both Contribution 2 (highlighting CFKC to address people-related vulnerabilities) and Contribution 3 (guiding frameworks to incorporate adaptive mechanisms for evolving human-centric challenges).
4. **RQ4: What process factors have evolved in the implementation and management of ZTA since the earlier studies?** Evolving compliance and governance challenges require new strategies. This question links to both Contribution 2 (addressing process-related disruptions due to generative AI) and Contribution 3 (proposing advanced governance strategies to strengthen ZTA frameworks).
5. **RQ5: How have technological factors, including advancements in cybersecurity tools, cloud environments, and automation, influenced ZTA implementation, and is the current ZTA knowledge base still relevant?** Technological advancements necessitate adaptive security frameworks. This question aligns with both Contribution 2 (highlighting technical vulnerabilities due to generative AI technologies) and Contribution 3 (identifying future technological directions for ZTA).

3. Classification of Existing Studies on ZTA

Unlike traditional security models that rely on perimeter defenses, ZTA operates on the principle of "never trust, always verify", ensuring that all access requests are continuously authenticated and authorized, regardless of their origin. This section aims to provide a comprehensive review and classification of existing studies on ZTA, focusing on both theoretical foundations and practical implementations. In addition to classifying studies by NIST SP 800-207 components, we explicitly annotate how each approach or deployment variant does—or does not—model generative-AI-enabled attacks and AI-based defences.

3.1. The Zero-Trust Architecture

ZTA is a cybersecurity framework that eliminates implicit trust within networks, enforcing continuous verification of users and devices to access resources [15]. The principle "never trust, always verify" forms the core of ZTA, ensuring that no entity can access sensitive resources without strict validation. ZTA emerged as traditional perimeter-based security models became inadequate due to increasingly sophisticated cyber threats, the rise of remote work, cloud computing, and mobile devices. The term "Zero Trust" was popularized by John Kindervag in 2010 [16], and Google operationalized such principles with its BeyondCorp initiative in 2014 [17], which emphasized secure access based on user identity and context, rather than network location. NIST SP 800-207, published in 2020, formalized ZTA by outlining its key components, including the Policy Engine (PE), Policy

Administrator (PA), and Policy Enforcement Point (PEP), enabling dynamic, granular access control. As shown in Figure 1, ZTA has evolved to address the complexities of modern IT environments and is widely adopted across sectors such as government, finance, and healthcare due to its focus on continuous verification, real-time threat monitoring, and reduced attack surfaces. The ZTA access control process, detailed in Algorithm A1 in the Appendix A, emphasizes continuous identity verification, contextual information analysis, and dynamic threat detection to allow or revoke access in real-time, improving resilience against modern cyber threats.

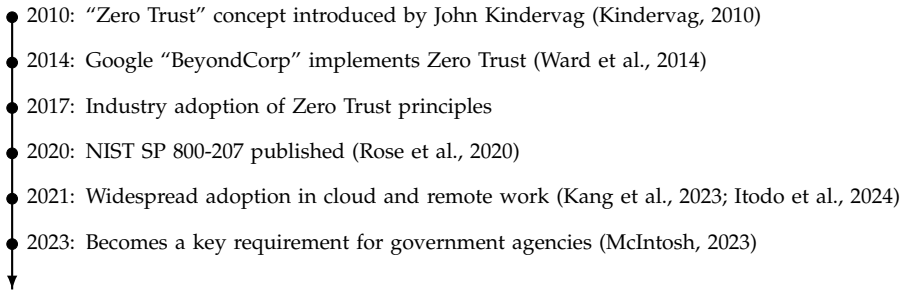

Figure 1. Timeline of key developments in zero trust architecture (ZTA). Sources: [4,8,12,16–18].

3.2. Core Components and Taxonomy of ZTA

This section elaborates on the ZTA core components and provides a structured taxonomy based on the NIST ZTA framework.

3.2.1. Variations of ZTA Approaches

Different approaches to implementing ZTA have emerged to address the diverse security needs of organizations. These variations reflect the flexibility and adaptability of ZTA in different contexts, highlighting the need for a tailored approach based on specific organizational requirements.

- *ZTA Using Enhanced Identity Governance:* Enhanced identity governance is crucial for maintaining strict control over who accesses what within an organization, and ensures that identity management systems are continuously verifying and re-verifying user and device credentials, which is critical in environments with high turnover, remote workforces, or frequent third-party access [19]. By emphasizing robust identity governance, organizations can reduce the risk of unauthorized access and ensure compliance with regulatory requirements.
- *ZTA Using Micro-Segmentation:* Micro-segmentation is a strategic approach that enhances security by dividing the network into smaller, isolated segments, which limits the potential damage that can be caused by a breach, as attackers are confined to a small segment of the network, rather than having free rein across the entire environment [20]. The granular control provided through micro-segmentation is particularly beneficial in cloud environments, where the traditional network perimeter is no longer as clearly defined.
- *ZTA Using Network Infrastructure and Software Defined Perimeters:* Modern network infrastructure, combined with Software Defined Perimeters (SDPs), offers a dynamic and scalable security solution that can adapt to the evolving threat landscape [21]. SDP technology allows for the creation of individualized, secure access pathways for each user, thereby reducing the attack surface. This approach is particularly effective in environments with a high degree of variability in access points, such as those involving IoT devices or remote workers.

3.2.2. Deployed Variations of the Abstract Architecture

The practical implementation of ZTA varies, depending on the specific needs and constraints of the organization. The following models represent common deployment variations, each offering unique advantages based on the operational environment.

- *Device Agent/Gateway-Based Deployment:* This deployment model involves installing agents or gateways on devices to enforce security policies and monitor activity in real time [22]. It is particularly effective in organizations where devices are highly mobile or where users frequently connect to the network from various locations. The centralized control provided by device agents ensures that security policies are uniformly applied across all devices, regardless of their location or status.
- *Enclave-Based Deployment:* Enclave-based deployments create isolated, secure zones within the network, which are ideal for protecting highly sensitive data or operations [23]. This model is often used in environments that require a high degree of separation between different departments or functions, such as government agencies or financial institutions. By creating enclaves, organizations can ensure that even if one part of the network is compromised, the rest remains secure.
- *Resource Portal-Based Deployment:* Resource portal-based deployment utilizes secure portals to control access to resources, providing a highly controlled entry point for users [24]. This model is particularly useful in environments where external partners or clients require access to specific resources without being granted broader network access. By channeling access through secure portals, organizations can maintain strict control over who accesses their most critical assets.
- *Device Application Sandboxing:* Sandboxing is a technique that isolates applications and processes from the rest of the network, preventing potential threats from spreading [25]. This approach is particularly valuable in environments where untrusted or third-party applications are frequently used. By isolating these applications, organizations can minimize the risk of malware or other threats compromising the network.

3.2.3. Trust Algorithm

Trust algorithms are the backbone of the decision-making process in ZTA. They continuously assess the trustworthiness of entities requesting access, ensuring that only those meeting the strict criteria are granted access [26]. The adaptability of trust algorithms allows organizations to fine-tune their security posture to address emerging threats or changes in the operational environment [27]. Below are common variations of trust algorithms, each tailored to specific security needs and operational contexts:

- *Risk-Based Trust Algorithms:* To dynamically adjust access decisions based on a calculated risk score [28]. The risk score is typically derived from factors such as user behavior patterns, device security posture, and the sensitivity of the requested resource. Risk-based algorithms enable organizations to implement granular access controls, where higher-risk actions or entities require additional verification steps or are denied access altogether.
- *Context-Aware Trust Algorithms:* To incorporate real-time contextual information, such as the geographic location of the access request, time of day, and the usual behavior of the user or device [29]. By analyzing these factors, context-aware algorithms can detect anomalies that might indicate a potential security threat, thereby enhancing the accuracy of access decisions.
- *Behavior-Based Trust Algorithms:* To focus on the continuous monitoring of user and entity behavior to establish a baseline of normal activity [30]. Any deviation from this baseline triggers a reassessment of trust. Behavior-based algorithms are particularly

effective in identifying insider threats or compromised credentials, as they can detect subtle changes in how users interact with systems.

- *Multi-Factor Trust Algorithms:* To combine multiple sources of information, such as biometrics, device health checks, and network conditions, to make comprehensive trust decisions [31]. Through integrating diverse factors, these algorithms provide a more robust and layered security approach, ensuring that access is granted only when all conditions meet the organization's security standards.
- *Adaptive Trust Algorithms:* To continuously evolve based on new data and threat intelligence [32]. They can adjust the criteria for trust dynamically, depending on the current threat landscape or changes in organizational policies. This variation is particularly valuable in environments with rapidly changing security requirements, as it allows the algorithm to "learn" from previous incidents and improve over time.

3.2.4. Network/Environment Components

The network and environment components are critical to supporting ZTA, providing the infrastructure necessary to accommodate its rigorous demands for continuous verification and dynamic access control. The following subcategories outline the key requirements and components needed to ensure a secure and efficient zero-trust environment:

- *Network Segmentation and Micro-Segmentation:* To limit lateral movement within the network, ZTA requires the implementation of network segmentation and micro-segmentation [33]. This involves dividing the network into smaller, isolated segments that can be individually monitored and controlled. Effective segmentation reduces the potential impact of a security breach by confining the threat to a specific segment, thereby protecting the broader network infrastructure.
- *Secure Communication Protocols:* Ensuring secure communication across all network components is essential for ZTA, and it includes the use of encrypted protocols such as TLS (Transport Layer Security) and IPsec (Internet Protocol Security) to protect data in transit [34]. These protocols help maintain the confidentiality and integrity of data exchanges between users, devices, and resources within the zero-trust environment.
- *High-Performance Authentication and Authorization Systems:* ZTA demands robust systems capable of handling large volumes of authentication and authorization requests in real time without introducing significant latency [35]. This includes implementing scalable identity management systems, such as federated identity and single sign-on (SSO) solutions, that can efficiently process and verify access requests while maintaining optimal performance.
- *Continuous Monitoring and Logging Infrastructure:* Continuous monitoring is a cornerstone of ZTA, requiring an infrastructure that can capture, analyze, and respond to security events in real-time [36]. This includes deploying security information and event management (SIEM) systems, intrusion detection systems (IDS), and advanced analytics platforms that provide comprehensive visibility into network activity and enable proactive threat detection and response.
- *Resilient and Redundant Network Architecture:* To support the continuous operation of ZTA, the network architecture must be resilient and capable of withstanding disruptions [37]. This involves implementing redundancy through failover mechanisms, load balancing, and distributed network resources to ensure that critical security functions remain operational even in the event of a failure or attack.
- *Integration with Cloud and Hybrid Environments:* Many organizations operate in cloud or hybrid environments, requiring ZTA to seamlessly integrate with these infrastructures [38]. This includes ensuring that zero-trust principles extend to cloud-based resources, with secure access controls, consistent policy enforcement, and visibility

across both on-premises and cloud environments. Proper integration ensures that the security posture is maintained, regardless of where data and applications reside.

3.3. Steps in Implementing ZTA

The implementation of ZTA involves a systematic approach that ensures continuous verification and dynamic access control, which are critical to maintaining a secure and resilient environment. The following structured steps demonstrate the essential phases in implementing ZTA, with a critical review of how existing studies have addressed such aspects.

3.3.1. Identifying Verification Triggers (When to Verify)

The first step in implementing ZTA is identifying the specific conditions and contexts that necessitate verification. Continuous verification is central to ZTA, as it dynamically assesses the trustworthiness of entities requesting access, ensuring that only authenticated and authorized entities are granted access to resources [1,39]. The existing literature has explored various scenarios where verification is required, such as during user authentication, device compliance checks, and network access requests. However, although existing studies have addressed verification timing and conditions, there is a scarcity of research on adaptive verification triggers. These triggers should be capable of accounting for real-time changes in user behavior, device health, and the evolving threat landscape, necessitating further investigation.

3.3.2. Verification Methods and Technologies (How to Verify)

The second step involves the selection and application of appropriate verification methods and technologies. ZTA relies on a variety of technologies, including identity and access management (IAM) systems, multi-factor authentication (MFA), and behavioral analytics, to ensure that every access request is rigorously authenticated and authorized. While existing research has proposed a range of verification mechanisms, the effectiveness of these methods often depends on the specific context in which they are applied. The current body of literature predominantly focuses on traditional verification methods, with limited exploration of emerging techniques such as AI-driven verification and continuous behavioral monitoring. There is a pressing need for further studies on integrating advanced verification technologies in dynamic and complex environments to enhance the effectiveness of ZTA.

3.3.3. Validation of Verification Processes (Verification Validation)

The final step in the implementation process is the validation of verification processes to ensure their reliability and accuracy. Proper validation is essential to maintaining the integrity of the ZTA framework. This involves cross-referencing the outcomes of verification against predefined security policies and employing audit logs and real-time monitoring to identify and address any discrepancies. While many studies advocate for various validation methods, such as the deployment of policy enforcement points (PEPs) and security information and event management (SIEM) systems, continuous real-time validation remains a significant challenge. Despite the critical importance of validation within ZTA, there is a notable lack of research on continuous validation techniques that can adapt to evolving threats and dynamic user behaviors. Developing robust validation frameworks that can maintain the reliability of ZTA in real time is an area that requires further exploration and innovation.

4. Literature Review Methodology

The literature review methodology was carefully designed to ensure a rigorous and comprehensive assessment of relevant studies, including systematic article selection, evaluation criteria, and a structured analysis approach.

4.1. Article Selection Criteria

The review sought studies that examine ZTA in a security context and that appeared after ZTA moved from policy discussion to mainstream deployment. The National Institute of Standards and Technology released *SP 800-207* [4] in late 2020; large-scale generative-AI tooling (for example, GPT-3.5 and Llama models) reached public availability since 2022. Work published before that date rarely considers the combined threat surface we analyze. For this reason, we restricted the sampling window to January 2022–August 2024, and our writing started in August 2024.

Search phase. We queried Google Scholar using the Boolean pattern

$$\text{“zero trust”} \lor \text{“zero-trust”} \lor \text{“ZTA”},$$

applied to title or author keywords. Google Scholar returned the union set with 978 unique records; the discipline-specific indexes added no new items once title-level duplication had been removed.

Inclusion criteria. A publication advanced to full-text screening when it satisfied *all* of the following:

1. Published in a peer-reviewed venue—journal special issue, conference proceedings, or magazine with documented editorial review;
2. Primary research, rather than a secondary survey;
3. Explicit focus on zero trust or a named ZTA component (e.g., policy engine, continuous verification);
4. Full text available in English.

Exclusion criteria. During screening, we excluded items that

1. Lacked an evaluation section or any empirical evidence;
2. Relied on unverifiable data (for instance, simulated traffic with no parameter disclosure);
3. Appeared in venues flagged by CABELLS or BEALL as predatory;
4. Were pre-prints, technical reports without peer review, or corporate white-papers.

Outcome. The filtering steps are summarized in Figure 2. Of 978 candidate papers, 368 met every inclusion rule. A further 231 were removed under the exclusion tests, leaving 136 primary studies, plus the 10 survey papers used as baselines. Then, we augmented our query with “generative AI” and synonyms (e.g., “large language model,” “diffusion model”) combined via AND with our ZTA terms to capture any emergent work at the intersection—even though we found virtually none.

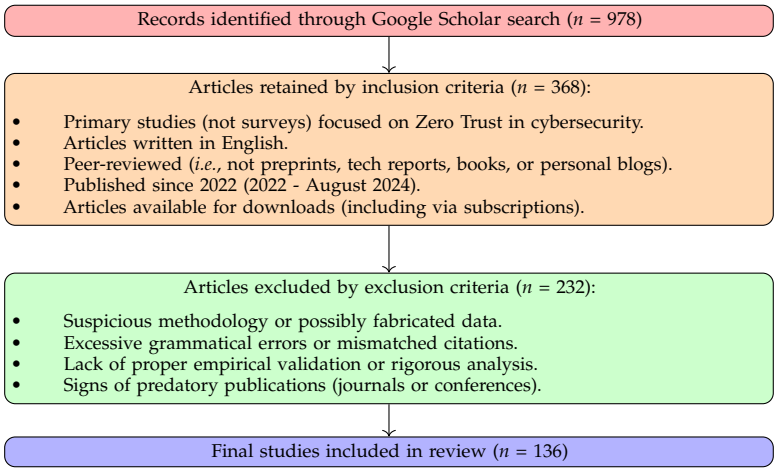

Figure 2. Flowchart of article selection process

4.2. Article Assessment Criteria

To evaluate the selected studies on ZTA, we employed a unified set of six evaluation criteria designed for primary research: academic rigor and scientific soundness, completeness of the three ZTA core principles, result replicability and code up-to-dateness, implementation versatility, practicality, and research ethics. The intent is to provide a consistent and objective framework for comparison across heterogeneous topics such as identity management, trust algorithms, network architecture, and deployment archetypes. Each criterion is scored on a three-level scale {Good = 1, Pass = 0.5, Fail = 0} using operationalized indicators with measurable thresholds and concrete examples, as detailed in Table 2. Unless otherwise stated, all six criteria are equally weighted for the main analysis, and the Total Score reported in Appendix B (Table A1) is the simple sum across criteria in $[0, 6]$ with no normalization. We additionally report a sensitivity check in Appendix B, where alternative weights are explored for robustness with no material change in study ordering.

Rater training and reliability. Two raters who are authors of this article independently scored all papers after a 90-minute calibration session that used five exemplar papers and a shared decision log. Disagreements were adjudicated by a third author who did not participate in the initial scoring. On a randomly selected calibration sample of ten papers not used in training, overall Krippendorff's $\alpha = 0.74$ (ordinal), with per-criterion α in the range $[0.66, 0.82]$, and raw percent agreement at 83% averaged across criteria. Full reliability tables, coder prompts, and the decision log template are included in Appendix B. These procedures aim to reduce subjectivity while keeping the rubric usable across diverse study designs.

Scoring and linkage to Appendix tables. The per-paper scores in Appendix B (Table A1) directly instantiate the rubric above, using the same criterion names, the same three-level scale, and equal weighting. When a study straddled categories, raters applied the decision log to maintain consistency, and any post-adjudication changes are marked in the log referenced in Appendix B. This kept the long table valid while making the rubric auditable and reproducible.

Table 2. Operationalized marking rubric for evaluating primary research on ZTA. Scoring is {Good = 1, Pass = 0.5, Fail = 0} with measurable indicators and examples.

Criteria	Good (1)	Pass (0.5)	Fail (0)
Academic rigor and scientific soundness	**Indicators:** Clear research question; methods justified and aligned to question; dataset or corpus described with sampling and inclusion criteria; statistical tests or formal models reported with effect sizes or confidence intervals where applicable; threats to validity enumerated. **Thresholds:** $\geq$3 of 4 documentation elements present (design, data, analysis, validity); formal analysis reproducibly specified or theoretically grounded. **Example:** An ablation study of a trust model with effect sizes and a validity section.	**Indicators:** Method description present yet partial; dataset mentioned but sampling or inclusion criteria missing; some analysis reported without effect sizes; high-level validity note. **Thresholds:** 2 of 4 documentation elements present. **Example:** Architecture paper with informal benchmarking and brief limitations.	**Indicators:** Objectives vague or absent; methods unclear; no analysis or unverifiable claims; no validity discussion. **Thresholds:** $\leq$1 of 4 documentation elements present. **Example:** Concept sketch with narrative assertions only.
Completeness of the three ZTA core principles ("Know when to verify", "How to verify", "Validate verification")	**Indicators:** All three principles specified and connected to control points; trigger logic for verification, verification mechanisms, and post-decision validation or monitoring described. **Thresholds:** All 3 principles covered with concrete mechanisms or policies. **Example:** A system defines risk triggers, FIDO2 step-up, and session-level post-decision checks.	**Indicators:** Two principles meaningfully covered; the third implicit or under-specified. **Thresholds:** 2 of 3 principles with at least one concrete mechanism. **Example:** Risk triggers and verification described but no post-decision validation.	**Indicators:** One or none of the principles addressed; no tie to enforcement points. **Thresholds:** $\leq$1 principle covered. **Example:** Policy narrative without triggers or validation path.
Result replicability and code up-to-dateness	**Indicators:** Public code or artifacts; pinned dependencies; seeds or configuration files; data access instructions or synthetic data generator; environment details. **Thresholds:** Code or artifacts publicly accessible and buildable; last update within 18 months of publication; successful re-run instructions. **Example:** Repository with Dockerfile and CI manifest that reproduces key tables.	**Indicators:** Partial artifacts or pseudo-code; external dependency list without pins; data available on request. **Thresholds:** Some artifacts exist but cannot reproduce all results; last update older than 18 months or missing pins. **Example:** Code shared yet requires manual fixes to run.	**Indicators:** No artifacts; proprietary data without a synthetic substitute; insufficient environment detail. **Thresholds:** Nothing usable to re-run analyses. **Example:** Closed implementation with unverifiable metrics.
Implementation versatility	**Indicators:** Demonstrates portability across at least two distinct environments or vendors; documents integration paths for SaaS/PaaS/on-prem; configurable policies. **Thresholds:** Evidence of cross-environment deployment or a neutral reference architecture with mappings. **Example:** Prototype validated on Kubernetes and an on-prem gateway with the same policy language.	**Indicators:** Runs in one environment with claims of portability; partial mapping to other stacks. **Thresholds:** Single-environment implementation with portability notes but no second deployment. **Example:** Cloud-only prototype with future on-prem plans.	**Indicators:** Hard-wired to a single stack; cannot adapt to other control planes; vendor-locked assumptions. **Thresholds:** No portability evidence or mapping. **Example:** Custom gateway with nonstandard policies and no adapters.
Practicality	**Indicators:** Reports end-to-end latency or throughput; resource use on target hardware; operator workload impact; deployment constraints; cost or scaling notes. **Thresholds:** At least two quantitative operational metrics (for example, p95 latency and CPU or memory) or a documented SLO; clear deployment prerequisites. **Example:** Access decision p95 $\leq$ 200 ms on commodity hardware with operator runbook.	**Indicators:** Qualitative claims of performance or usability; one quantitative metric without context. **Thresholds:** Some indication of feasibility without full operational picture. **Example:** "Low overhead" claim with a single median latency value.	**Indicators:** No operational metrics; ignores operator or cost considerations; cannot infer feasibility. **Thresholds:** No quantitative practicality evidence. **Example:** Algorithm-only study without deployment envelope.
Research ethics	**Indicators:** Self-disclosed limitations; bias and threat discussion; dataset licensing and consent statements where human data are involved; security disclosure posture; clear future work. **Thresholds:** Explicit limitations and bias section plus data use or disclosure statements when applicable. **Example:** Paper documents sampling bias and provides an IRB or equivalent statement.	**Indicators:** Mentions some limitations or bias; omits data handling details or disclosure posture. **Thresholds:** Partial coverage of ethical dimensions. **Example:** Limitations paragraph without data licensing notes.	**Indicators:** No discussion of limitations, bias, or data use; no disclosure posture; unsafe release practices. **Thresholds:** Ethical aspects absent. **Example:** Uses scraped personal data with no consent or licensing.

5. Evaluation of Existing Primary Studies

This section systematically evaluates 136 existing primary studies on ZTA during 2022–August 2024, categorizing them based on different architectural variations and assessing their practical applicability, scalability, and alignment with zero-trust principles (Table A1 in the Appendix B).

5.1. Variations of ZTA Approaches

As indicated in Algorithm 1, studies under this subsection exhibit a common procedural flow in developing ZTA approaches. This process typically initiates with the identification of security challenges, integrates various technologies such as blockchain or AI, and culminates in the proposal of ZTA models. A critical observation is that these studies frequently lack real-world scalability and validation, indicating a convergence on theoretical frameworks with a divergence in practical applicability.

Algorithm 1 Summarized common structure of ZTA variation studies

Input: Security environment $\mathcal{E}$, emerging technologies $\mathcal{T}$ (e.g., Blockchain, AI)
Output: Proposed ZTA implementation $\mathcal{Z}$
Identify Security Challenges $\mathcal{C}$ within $\mathcal{E}$ ▷ Start
Integrate Technologies $\mathcal{T}$ into $\mathcal{C}$ ▷ Blockchain, AI
Develop ZTA Model $\mathcal{M}$ ▷ Abstract architecture design
Propose ZTA Implementation $\mathcal{I}(\mathcal{M}, \mathcal{T})$
Address Scalability $\mathcal{S}$ issues $\Rightarrow \mathcal{S}(\mathcal{I}) \approx \mathcal{I}$
Evaluate Real-World Validation $\mathcal{V}(\mathcal{I})$ ▷ Validation in practical scenarios
Analyze Gaps $\mathcal{G}(\mathcal{I})$ ▷ End

5.1.1. ZTA Using Enhanced Identity Governance

Five studies [26,40–43] explored the implementation of ZTA with a focus on enhanced identity governance. Ref. [40] developed a ZTA tailored for health information systems, emphasizing strict identity governance to enhance security; however, the study lacked a comprehensive analysis of how scalable the implementation would be across diverse healthcare environments. Ref. [41] proposed a blockchain-based approach to reinforce multi-factor authentication within ZTA to secure digital identities, but it did not provide a detailed discussion on scalability and integration with existing systems. Ref. [42] examined ZTA's application to digital privacy in healthcare, focusing on enhanced identity governance; however, it inadequately addressed the adaptability of ZTA in dynamic healthcare settings. Ref. [26] implemented a ZTA approach aimed at enhancing information security for healthcare workers through digital health technology adoption, but the study fell short in addressing the specific challenges associated with dynamic and context-aware access controls. Ref. [43] applied ZTA principles, along with probability-based authentication, to improve data security and privacy in cloud environments, but it did not include a comprehensive evaluation of the scalability of this approach across different cloud platforms. While these studies contributed valuable insights into the role of identity governance in ZTA, they collectively overlooked detailed evaluations of scalability and adaptability, which are crucial for the practical deployment of ZTA in various real-world scenarios.

5.1.2. ZTA Using Micro-Segmentation

Five studies [44–48] explored the implementation of ZTA with an emphasis on micro-segmentation. Ref. [44] developed a zero-trust machine learning architecture specifically for healthcare IoT cybersecurity, highlighting the importance of green computing. However, the study lacked a comprehensive evaluation of the practical challenges associated with deployment in real-world scenarios. Ref. [45] proposed a methodology for auto-

mated zero-shot data tagging using generative AI to facilitate tactical ZTA implementation, yet it did not adequately address scalability issues in diverse operational environments. Ref. [46] introduced a ZTA approach tailored to cloud-based fintech services, emphasizing micro-segmentation to enhance security, but the study fell short of providing a thorough analysis of the limitations inherent in the proposed approach. Ref. [47] investigated the application of Zero-Trust principles in microservice-based smart grid operational technology (OT) systems to mitigate insider threats, but it lacked a detailed evaluation of scalability, particularly in large-scale environments. Ref. [48] presented a framework for applying zero-trust principles to Docker containers, focusing on micro-segmentation within containerized environments; however, it did not include a detailed analysis of the framework's performance across varied operational settings. In summary, while such studies reveal the significance of micro-segmentation in enhancing ZTA, they collectively fall short in addressing practical deployment challenges and scalability, which are essential for implementing ZTA effectively across different environments.

5.1.3. ZTA Using Network Infrastructure and Software Defined Perimeters

Sixty-eight studies [13,40,49–114] explored the integration of ZTA using network infrastructure and Software Defined Perimeters (SDP). Refs. [49–52] focused on leveraging SDPs to enhance ZTA deployment in virtual, cloud, and metaverse environments, highlighting the potential of blockchain and AI integration for dynamic and scalable security. However, they lacked a thorough analysis of real-world deployment challenges, which is critical for practical applicability. Refs. [13,53,55,60] examined ZTA implementations in IoT and mobile networks, emphasizing the role of network infrastructure and SDPs to enhance security. These studies identified key differences in ZTA deployment strategies but often fell short of addressing the integration challenges with legacy systems and scalability in large, heterogeneous environments. Similarly, Refs. [54,56,59] proposed decentralized models for ZTA using blockchain and per-packet authorization, yet they lacked comprehensive validation in diverse real-world scenarios, limiting the generalizability of their findings. Studies like [61,65,67] focused on enhancing ZTA in IoV and connected vehicle settings through distributed edge solutions and federated learning. However, their research often did not extend to broader trust validation processes beyond specific use cases like caching or intrusion detection. Meanwhile, Refs. [64,68,69] explored innovative methods like quantum computing and reputation-enhanced schemes to bolster ZTA, but again, the absence of real-world applicability evaluations was a significant gap. Refs. [71,72,75] addressed ZTA in specialized environments such as 6G networks and mandatory access control frameworks, but they inadequately covered practical implementation details. Other studies, such as [73,74,78,81], reviewed modern network infrastructure applications for ZTA but lacked comprehensive analysis of hybrid deployment challenges. Furthermore, Refs. [79,80,88,96] examined ZTA in the context of cloud-native and industrial IoT settings. They focused on enhancing security through modern infrastructure but did not sufficiently address the scalability and integration issues that arise in complex, real-world IoT environments. Studies like [82,85,86] contributed to understanding the applicability of ZTA in edge computing and cyber-physical systems, yet their empirical validation remains a critical concern. Lastly, Refs. [40,81,87] provided innovative approaches to ZTA in healthcare or connected vehicle environments, using SDPs and dynamic infrastructure to secure access and communication. However, they too lacked comprehensive real-world deployment evaluations, which are essential for validating their proposed solutions. While such studies collectively highlight the importance of integrating network infrastructure and SDPs to implement ZTA across various domains, the recurring gaps in addressing practical de-

ployment challenges, scalability, and real-world validation limit their applicability and effectiveness in diverse operational settings.

5.2. Deployed Variations of the Abstract Architecture

As captured in Algorithm 2, studies in this subsection share a generic procedural structure for exploring deployment variations of ZTA. The approach typically begins by identifying deployment challenges, followed by proposing deployment models such as device agent, enclave-based, or resource portal-based strategies. The process then involves addressing scalability issues and evaluating real-world applicability. These studies often culminate in analyzing gaps in deployment models, with a common observation being the lack of practical validation in large-scale environments.

Algorithm 2 Summarized common structure in deployed variations of ZTA studies

1: **Input:** Deployment environment $\mathcal{E}$, Deployment models $\mathcal{D}$ (e.g., Device Agent, Enclave, Resource Portal)
2: **Output:** Evaluated ZTA deployment $\mathcal{Z}$
3: Identify Deployment Challenges $\mathcal{C}$ within $\mathcal{E}$ ▷ Start
4: Propose Deployment Model $\mathcal{M} \in \mathcal{D}$
5: Address Scalability $\mathcal{S}(\mathcal{M}) \rightarrow$ Modify $\mathcal{M}$ if $\mathcal{S}(\mathcal{M}) \not\approx \mathcal{E}$
6: Evaluate Real-World Applicability $\mathcal{V}(\mathcal{M}) \rightarrow$ Test $\mathcal{M}$ in $\mathcal{E}$
7: Analyze Gaps $\mathcal{G}(\mathcal{M})$ ▷ End

5.2.1. Device Agent/Gateway-Based Deployment

Two studies [115,116] explored the application of device agent and gateway-based deployment strategies within ZTA. Ref. [115] employed a device profiling method combined with deep reinforcement learning to detect anomalies, aiming to enhance the security of ZTA networks. However, this study inadequately addressed scalability issues, limiting its applicability across diverse and large-scale network environments. Meanwhile, Ref. [116] proposed a zero-trust model tailored to intrusion detection in drone networks using device agent-based deployment, but it lacked comprehensive validation, making its practical applicability in real-world scenarios uncertain. While both studies contributed valuable insights into the role of device agent/gateway-based deployment in ZTA, they did not fully explore scalability and real-world applicability, which are critical for the successful implementation of such models in dynamic environments.

5.2.2. Enclave-Based Deployment

No studies were found on enclave-based deployment, despite the critical role enclaves play in protecting sensitive data by isolating it from the rest of the system [15]. With the rise of hardware-based trusted execution environments (TEEs) like Intel SGX and AMD SEV, using enclaves can significantly strengthen ZTA by ensuring that, even if other security measures are compromised, sensitive data remains secure [117]. This gap in research leaves ZTA implementations vulnerable to sophisticated attacks that could exploit such unprotected areas, highlighting an urgent need for focused study and development in this area.

5.2.3. Resource Portal-Based Deployment

Despite its potential to enhance security by providing controlled access to critical resources, no studies were found that specifically address Resource Portal-Based Deployment within the context of ZTA. This lack of research represents a significant oversight, as secure resource portals can offer a robust solution for managing external access, especially in scenarios involving partners and clients who require limited, monitored access [15,118]. Given

the increasing need to safeguard sensitive data while allowing selective external access, more research should be dedicated to developing and evaluating resource portal-based deployment strategies in ZTA to fill this critical gap.

5.2.4. Device Application Sandboxing

Ref. [119] was the only study found in this category. While the study proposed a hybrid isolation model for device application sandboxing within ZTA, it lacked comprehensive real-world implementation and performance evaluations. Device application sandboxing is essential for ZTA, as it isolates applications to prevent malicious code from affecting other parts of the system, thereby enhancing security [15]; however, it appears to be an understudied area since 2022, which is concerning, given the increasing sophistication of cyber threats that exploit application vulnerabilities, leaving a critical gap in ZTA research and implementation strategies.

5.3. Trust Algorithm

The studies under this subsection follow a generalized workflow when developing and evaluating trust algorithms within the ZTA. As detailed in Algorithm 3, the process typically initiates with identifying trust and security challenges and culminates in deployment and monitoring in real-world scenarios. The iterative nature of trust algorithm development involves stages of algorithm development, simulation, optimization, validation, integration, and scalability evaluation. This abstract representation encapsulates the shared procedural elements across the studies in this category.

Algorithm 3 Summarized common structure in trust algorithm development and integration in ZTA

Input: Security Challenges $\mathcal{C}$, ZTA Framework $\mathcal{Z}$, Trust Algorithm $\mathcal{T}$
Output: Optimized and Deployed Trust Algorithm $\mathcal{T}^*$
Identify Security Challenges $\mathcal{C}$ ▷ Start
Develop Initial Trust Algorithm $\mathcal{T}$ ▷ Propose Model/Approach
Simulate $\mathcal{T}$ and Conduct Preliminary Testing $\mathcal{P}(\mathcal{T})$
Optimize $\mathcal{T} \Rightarrow \mathcal{T}^*$ via Parameter Tuning θ^*
Validate $\mathcal{T}^*$ through Experimental Evaluation $\mathcal{V}(\mathcal{T}^*)$
Integrate $\mathcal{T}^*$ into ZTA Framework $\mathcal{Z} \oplus \mathcal{T}^*$
Evaluate Scalability $\mathcal{S}(\mathcal{T}^*)$ and Real-World Applicability $\mathcal{R}(\mathcal{T}^*)$
Address Feedback $\mathcal{F}(\mathcal{T}^*)$ and Refine if Necessary
Deploy and Monitor $\mathcal{T}^*$ in Real-World Scenarios ▷ End

5.3.1. Risk-Based Trust Algorithms

Fourteen studies [120–133] explored various risk-based trust algorithms within the framework of ZTA. Of these, Refs. [120,122,123] primarily focused on enhancing decision-making processes by integrating risk assessment models but lacked concrete real-world deployment examples to validate their proposals. Studies like Refs. [121,127] proposed the use of blockchain and federated learning to manage trust dynamically in network environments, yet they inadequately addressed the scalability of these solutions in practical applications. Refs. [125,126,129] concentrated on developing frameworks to assess the maturity and trustworthiness of systems under ZTA, employing various analytical and algorithmic approaches. However, their limitations lay in the insufficient evaluation of these frameworks against diverse operational scenarios. On the other hand, Refs. [124,128] aimed to tackle specific challenges, such as those in 6G networks and intrusion detection, by formulating risk-based models but did not provide comprehensive details on practical implementation. Studies like Refs. [130,133] introduced innovative risk-scoring algorithms tailored to endpoint devices and network access control, focusing on dynamic adjustments

based on security profiles. Despite their innovative approaches, they lacked a thorough analysis of the deployment challenges in real-world settings. Finally, Refs. [131,132] explored user behavior risks and zero-trust data management but failed to address the full spectrum of implementation constraints, leaving gaps in their practical relevance.

5.3.2. Context-Aware Trust Algorithms

Five studies [134–138] explored the development and implementation of context-aware trust algorithms within the framework of ZTA. Ref. [134] implemented a dual fuzzy methodology to enhance trust-aware authentication and task offloading in multi-access edge computing. However, the study lacked a detailed discussion on the challenges associated with real-world deployment, leaving a gap in understanding its practical applicability. Ref. [135] examined the influence of information security culture on the adoption of ZTA within UAE organizations. While the study provided valuable insights into cultural impacts on security adoption, its primary shortcoming was its limited generalizability beyond the specific cultural context of the UAE, making its findings less applicable to broader or more diverse environments. Ref. [136] introduced the ZETA framework, which integrates split learning with zero-trust principles to enhance security in autonomous vehicles operating within 6G networks. Despite its innovative approach, the study did not provide a comprehensive evaluation of the practical implementation and scalability of the framework in real-world settings, leaving unanswered questions about its feasibility and effectiveness under varied operational conditions. Ref. [137] proposed a continuous authentication protocol designed for ZTA that operates without a centralized trust authority, but it did not undergo thorough testing for scalability in real-world environments, which limits the ability to gauge its performance under different network loads and conditions. Ref. [138] presented a context and risk-aware access control approach tailored for zero-trust systems. The study contributed significantly to the theoretical development of context-aware access controls but fell short in evaluating the practical implementation of its approach across diverse operational environments, which is crucial for understanding the broader applicability and effectiveness of the proposed solutions. While these studies have made significant strides in developing context-aware trust algorithms within ZTA, they share common limitations in terms of real-world deployment, scalability, and generalizability, which need to be addressed in future research to fully realize the potential of these innovations.

5.3.3. Behavior-Based Trust Algorithms

No studies were identified that specifically focused on behavior-based trust algorithms within the context of ZTA. This gap is concerning, given the increasing sophistication of modern cyber threats, such as advanced persistent threats (APTs) and social engineering attacks, which often involve compromised insider credentials [15]. Traditional risk-based and static trust algorithms cannot adequately detect these threats, as they rely on predefined rules or known threat patterns [5]. In contrast, behavior-based algorithms excel at identifying anomalies by continuously monitoring user and entity behavior against established baselines. The absence of research in this area leaves a critical vulnerability in ZTA implementations, as these algorithms are uniquely capable of detecting subtle deviations indicative of insider threats or compromised accounts, which are otherwise challenging to identify with conventional methods.

5.3.4. Multi-Factor Trust Algorithms

Two studies [139,140] focused on the development and implementation of multi-factor trust algorithms within the context of ZTA. In [139], the authors explored the automation and orchestration of ZTA, with a particular emphasis on the challenges associated with developing trust algorithms. While this study provided valuable insights into the theoretical

aspects of algorithm design, it fell short in terms of comprehensive real-world applicability testing, leaving questions about the practical deployment of such algorithms unanswered. In contrast, Ref. [140] proposed a data-driven zero-trust key algorithm aimed at enhancing security by dynamically adjusting trust levels based on multiple data inputs. This approach suggested a promising method for improving the responsiveness and accuracy of trust decisions in dynamic environments. However, similar to [139], the study by Liu et al. lacked a thorough evaluation of the real-world implementation challenges. The authors did not address how these algorithms would perform under the constraints of existing IT infrastructures or in the presence of advanced persistent threats (APTs). While both studies contributed to the academic discourse on multi-factor trust algorithms in ZTA, they were limited by a lack of empirical testing and practical validation. This gap in real-world applicability requires future research that bridges the divide between theoretical algorithm design and practical implementation.

5.3.5. Adaptive Trust Algorithms

Six studies [26,39,141–144] were found to have explored adaptive trust algorithms within ZTA. Of these, Ref. [141] proposed a transition from standard policy-based zero trust to an absolute zero-trust (AZT) model utilizing quantum-resistant technologies. While innovative, this study lacked comprehensive practical implementation and real-world validation, limiting its applicability in real-world scenarios. In [142], the authors implemented a blockchain-based zero-trust approach integrated with deep reinforcement learning specifically for supply chain security. However, this study inadequately addressed the practical deployment challenges and the energy consumption of the proposed solution, raising concerns about its feasibility. Ref. [39] critically analyzed various trust algorithms within ZTA, emphasizing the importance of continuous verification. Despite its detailed analysis, the study fell short in evaluating the implementation across real-world deployment scenarios, leaving a significant gap in practical applicability. Similarly, Ref. [143] explored continuous authentication methods within a zero-trust framework, focusing on the potential of adaptive algorithms to enhance security. However, the study's major inadequacy was the limited practical implementation details and the absence of comprehensive performance evaluation metrics, which are crucial for assessing real-world effectiveness. In [144], an AI-based approach was proposed for implementing ZTA, focusing on the use of adaptive trust algorithms to improve security decision-making processes. However, the study lacked comprehensive evaluation across multiple deployment environments, which limits its generalizability and raises concerns about its adaptability. Finally, Ref. [26] developed a maturity framework for zero-trust security in multi-access edge computing. Although the framework was well conceived, it inadequately addressed the versatility of implementation across different platforms, which is essential for a robust zero-trust deployment. While these studies explored various adaptive trust algorithms that could significantly advance zero-trust principles, most failed to thoroughly address the practical implementation challenges and performance metrics necessary for real-world deployment.

5.4. Network/Environment Components

The studies under this subsection demonstrate a recurring pattern in the lifecycle of network/environment components within ZTA. As detailed in Algorithm 4, the process typically involves identifying operational needs, followed by development, installation, and validation of components. This generic workflow emphasizes the iterative nature of integration, monitoring, and feedback within real-world deployments, highlighting the shared approach across various studies.

Algorithm 4 Summarized common lifecycle of network/environment components in ZTA

Input: Operational Needs $\mathcal{O}$, Component $\mathcal{C}$
Output: Integrated and Monitored Component $\mathcal{C}^*$
Identify Operational Needs $\mathcal{O}$
Develop or Procure Component $\mathcal{C}$
Install and Configure $\mathcal{C}$
Validate and Test $\mathcal{C}$ ▷ Validation: $\mathcal{V}(\mathcal{C})$
Integrate $\mathcal{C}$ with Existing Systems $\mathcal{I}(\mathcal{C})$
Deploy $\mathcal{C}$ in Production Environment
Continuous Monitoring and Feedback $\mathcal{M}(\mathcal{C})$
if Feedback $\mathcal{F}(\mathcal{C})$ Indicates Revisions **then**
    Iterate: Return to Step 5
end if
Final Integration of Component $\mathcal{C}^*$

5.4.1. Network Segmentation and Micro-Segmentation

Twelve studies [145–156] examined various aspects of network segmentation and micro-segmentation within ZTA. Among them, Refs. [145,146] focused on integrating hardware-level monitoring and zero-trust principles to secure zero-trust networks and cyber–physical systems. However, neither study extensively evaluated their adaptability across diverse environments. Studies like [147,148] developed software integrity protocols and reviewed security in the metaverse, respectively, but failed to provide detailed implementation strategies. While Refs. [149,151] leveraged machine learning and AI to enhance micro-segmentation and adaptive authentication, they did not sufficiently address real-world deployment challenges. Other studies, such as [150,153,154], explored de-perimeterization and graph-based pipelines for zero trust, yet they lacked practical validation and comprehensive guidelines for implementation. Finally, Refs. [155,156] applied micro-segmentation to secure the Internet of Vehicles and digital forensics but did not thoroughly evaluate their applicability and versatility in diverse operational settings.

5.4.2. Secure Communication Protocols

Seven studies [3,34,157–161] were identified that explored secure communication protocols within ZTA. Ref. [157] analyzed the intersection of intellectual property rights with information security technologies, touching on secure communication protocols. However, it inadequately addressed the practical implications of these technologies within operational environments, lacking an in-depth analysis of how these protocols would function under real-world constraints. Ref. [158] proposed integrating federated learning with a zero-trust approach to enhance security in wireless communications. While innovative, the study fell short in comprehensively addressing the complexities of applying this integration across varied real-world scenarios, particularly in environments with diverse network conditions and infrastructure. Ref. [159] introduced a privacy-preserving authentication scheme based on ZTA, emphasizing secure communication protocols. Despite the novel approach, the study did not sufficiently explore the practical implementation and scalability of the proposed solution in diverse environments, leaving gaps in understanding how these protocols would perform under different operational contexts. Ref. [3] investigated the use of proxy smart contracts to enhance ZTA within decentralized oracle networks. Although the study provided valuable insights into the integration of blockchain with ZTA, it lacked a thorough analysis of practical deployment scenarios and real-world performance metrics, making it difficult to assess the effectiveness of the proposed solution. Ref. [160] presented a defense model integrating zero-trust principles with blockchain technology to secure smart electric vehicle chargers. The study was forward-looking but failed to provide a detailed analysis of scalability and performance under varying network conditions, which

are critical for assessing the real-world applicability of the proposed protocols. Ref. [161] implemented a blockchain-based mechanism for securely delivering enrollment tokens in zero-trust networks. While the study contributed to the field by addressing secure token delivery, it lacked a comprehensive evaluation of its performance across different real-world scenarios, limiting its practical relevance. Ref. [34] proposed a framework for sustained zero-trust security in mobile core networks for 5G and beyond, focusing on secure communication protocols, but it did not include an in-depth analysis of the practical implementation challenges, such as the impact on network latency and the adaptability of these protocols in dynamic mobile environments. While these studies provide valuable contributions to the development of secure communication protocols within ZTA, most fall short in addressing the practical challenges of real-world deployment, including scalability, performance metrics, and adaptability to diverse operational environments.

5.4.3. High-Performance Authentication and Authorization Systems

This subsection reviews three studies [157,162,163] that focus on the implementation and challenges of high-performance authentication and authorization systems within ZTA. Ref. [157] analyzed the intersection of intellectual property rights with information security technologies. While the study highlighted the importance of safeguarding intellectual property within secure environments, it inadequately addressed the practical implications of these technologies in operational settings, particularly how these systems could be deployed and maintained in real-world scenarios. Ref. [162] proposed a zero-trust decentralized mobile network user plane by implementing dNextG to enhance security through decentralized access control mechanisms. Although the study provided a robust framework for improving security in mobile networks, it fell short in discussing the scalability challenges that arise in diverse and expansive network environments. The study did not adequately consider how the proposed solution would perform across different network scales, particularly in heterogeneous or large-scale deployments. Ref. [163] conducted an analysis of the cost-effectiveness of implementing ZTA in various organizations. While the study offered valuable insights into the financial aspects of ZTA deployment, it lacked a detailed discussion on adapting the proposed solutions to different organizational contexts. The research did not explore how varying organizational structures, resources, and operational needs might impact the effectiveness and adaptability of ZTA solutions. While these studies provide important contributions to the field of high-performance authentication and authorization systems within ZTAs, they each exhibit certain limitations. The practical deployment implications, scalability challenges, and contextual adaptability of the proposed solutions were often inadequately addressed, highlighting areas for further research and improvement.

5.4.4. Continuous Monitoring and Logging Infrastructure

No studies specifically focusing on continuous monitoring and logging infrastructure within the ZTA framework were found. This gap is critical, as continuous monitoring not only enables the early detection of threats by identifying anomalies in real time but also provides the capability to mitigate attacks before they escalate [15]. Logging, particularly with write-once, read-only configurations, ensures accountability and preserves forensic evidence, which is invaluable for post-incident analysis and compliance [15,30]. Ignoring these components leaves ZTA implementations vulnerable to sophisticated attacks that exploit the absence of continuous oversight and reliable audit trails.

5.4.5. Resilient and Redundant Network Architecture

Ref. [39] was the only study identified in this category. The study critically analyzed the infrastructure components required to support ZTA, emphasizing the importance of re-

silient and redundant network designs to ensure continuous availability and robust security. However, it lacked a thorough examination of practical implementation challenges and did not provide detailed evaluations of the architecture's adaptability in real-world scenarios. Resilient and redundant network architecture is vital for ZTA as it ensures that network disruptions or failures do not compromise security [15]; nonetheless, this area remains underexplored, especially concerning the practical integration of such architectures in diverse operational environments. The limited focus on implementation adaptability in [39] highlights a significant gap in current ZTA research, necessitating further investigation to address these challenges effectively.

5.4.6. Integration with Cloud and Hybrid Environments

Six studies [28,38,164–167] explored the integration of ZTA with cloud and hybrid environments, each addressing different aspects of this critical area. Ref. [38] developed a flexible ZTA tailored for the cybersecurity of industrial IoT infrastructures. However, the study did not include a comprehensive evaluation of the architecture's scalability and performance in diverse real-world settings, leaving questions about its practical applicability. Ref. [164] proposed a comprehensive framework for migrating to ZTA with a focus on network and environment components. Despite its thorough design, the study's major shortcoming was the limited empirical validation of the proposed framework, making it difficult to assess its effectiveness in practice. Ref. [28] investigated the feasibility of applying ZTA to secure the metaverse. While the study provided valuable insights into the potential benefits and challenges, it lacked an in-depth analysis of the performance impacts and practical challenges associated with its deployment in this emerging digital environment. Ref. [165] examined the potential integration of ZTA with emerging 6G technologies, highlighting various opportunities and challenges. However, the study was limited due to its lack of detailed analysis regarding specific implementation strategies, leaving a gap in understanding how ZTA could be effectively deployed within 6G networks. Ref. [166] explored secure and scalable cross-domain data sharing in a zero-trust cloud-edge-end environment using sharding blockchain. Although the study presented a novel approach, it did not sufficiently discuss the practical implications of sharding blockchain in real-world ZTA scenarios, particularly in terms of scalability and security. Ref. [167] augmented zero-trust network architecture to enhance security in virtual power plants. The study, while innovative, was constrained due to its limited focus on scalability in highly dynamic environments, which is a crucial factor in the successful deployment of ZTA in such settings. While these studies contribute to the growing body of knowledge on integrating ZTA with cloud and hybrid environments, they each exhibit significant gaps, particularly in empirical validation, scalability assessment, and the practical challenges of real-world implementation.

5.5. Major Themes Identified

Our evaluation of the primary studies revealed several overarching themes.

5.5.1. Overstatement of Research Success

A significant number of studies claim to have comprehensively addressed the challenges of ZTA. However, a deeper examination reveals a stark contrast between these claims and the actual rigor of their methodologies and adherence to critical ZTA principles. Among the 136 studies evaluated, the criterion for "academic rigor" shows a somewhat balanced distribution with 85 occurrences of a full score (1) and 52 occurrences of a partial score (0.5). This suggests that, while many studies demonstrate strong theoretical foundations, there is still a notable portion that lacks comprehensive methodological depth. More concerning is the evaluation against the "ZTA 3-Step Completeness" criterion, which is

fundamental according to the NIST ZTA guideline. In this area, only 2 studies achieved a full score (1), while a staggering 132 studies received a partial score (0.5), and 3 studies failed entirely (score of 0). This indicates that 98.54% of the studies did not fully meet the critical requirement of ZTA 3-Step Completeness mandated by NIST ZTA [15]. Despite these shortcomings, the studies often present an overly optimistic narrative, downplaying these gaps and overemphasizing their contributions.

5.5.2. Mixed Quality in Research Applicability, Versatility, and Practicality

The quality of research concerning applicability, versatility, practicality, and research ethics varied significantly. This is evident from the diverse scores assigned across the studies. For instance, within the *Network Infrastructure and Software Defined Perimeters* category, scores for practicality and versatility ranged from as high as 1 to as low as 0, indicating inconsistencies in how well these studies can be applied in real-world scenarios. The average score in this category for practicality was approximately 0.5, which suggests that many studies struggle with implementing their proposed solutions in a way that is both versatile and practical across different environments. This variability highlights the need for more rigorous and contextually adaptable research that addresses the broader challenges of implementing ZTA in diverse operational settings.

5.5.3. Selective Coverage of ZTA Topics

A critical observation is the uneven focus on certain aspects of ZTA in the existing research. Topics like *Network Segmentation and Micro-Segmentation* and *Network Infrastructure and Software Defined Perimeters* have been extensively explored, with 12 and 68 studies, respectively, indicating a strong emphasis on network isolation and infrastructure protection. However, significant areas such as *Enclave-Based Deployment*, *Resource Portal-Based Deployment*, *Behavior-Based Trust Algorithms*, and *Continuous Monitoring and Logging Infrastructure* are notably absent from the literature. This selective attention to specific ZTA components leaves critical security mechanisms underrepresented, which is concerning, given the complex and evolving nature of cyber threats. *Enclave-Based Deployment* could provide robust isolation for sensitive data but remains unexplored, limiting the understanding of its implementation and effectiveness. Similarly, *Behavior-Based Trust Algorithms*, essential for detecting sophisticated insider threats and compromised credentials, are missing, potentially leaving organizations vulnerable to undetected malicious activities. Furthermore, the absence of studies on *Continuous Monitoring and Logging Infrastructure* neglects the need for real-time threat detection and accountability through forensic evidence, which are fundamental for proactive security management and post-incident analysis.

As depicted in Figure 3, the majority of studies focused on *ZTA Approaches*, comprising 57% of the total. In contrast, only 2% cover *Deployed Variations*, while 20% and 21% focus on *Trust Algorithms* and *Network Components*, respectively. This major imbalance highlighted a pronounced research preference towards high-level architectural concepts over practical deployment and comprehensive trust mechanisms, showing the need for a more balanced approach that should include underrepresented yet critical areas to provide a holistic understanding of ZTA implementations.

5.6. Addressing Research Questions RQ1 and RQ2

The evaluation of existing primary research on ZTA reveals mixed success in addressing the first two research questions (RQ1 and RQ2), while the latter three (RQ3, RQ4, and RQ5) remained unanswered.

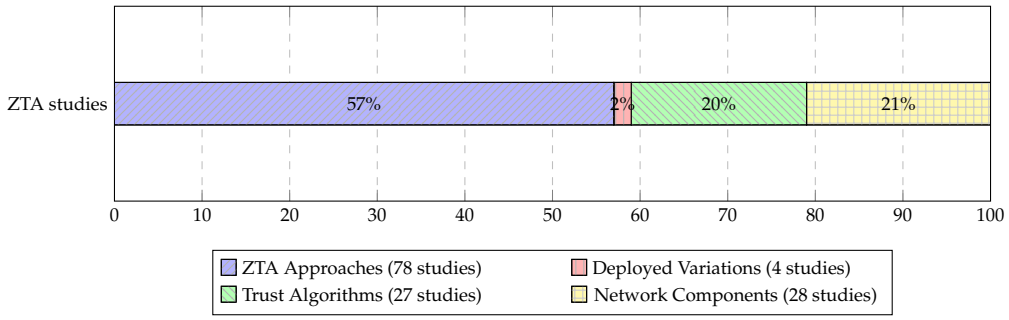

Figure 3. Percentage distribution of studies across major ZTA taxonomy classifications.

RQ1: How can "trust" and "zero trust" be properly defined within the context of ZTA? ANSWERED (although varied in quality). The surveyed studies have explored various theoretical underpinnings of trust and zero trust within ZTA. Several frameworks attempt to define trust as a dynamic, context-sensitive construct, where every interaction must be continuously verified. These definitions have laid the groundwork for understanding trust as a minimal and temporary concept within ZTA, which is fundamental for its implementation. However, the depth and clarity of these definitions differ across studies, with some failing to provide practical guidelines for operationalizing zero trust in diverse environments. Thus, while the question of how to define trust and zero trust has been addressed, the variation in quality and practical applicability limits the overall effectiveness of the existing answers.

RQ2: What are the different ways to achieve effective implementation of ZTA across diverse operational environments? ANSWERED (although varied in quality). The strategies for implementing ZTA across various environments have been examined, with a particular focus on risk-based access controls, network segmentation, and identity governance. Studies have outlined different methods to achieve ZTA, emphasizing the integration of continuous authentication, policy enforcement, and micro-segmentation as key mechanisms. However, the primary weakness lies in the lack of empirical testing in highly varied environments, such as multi-cloud or hybrid infrastructures. While the theoretical models provide a broad understanding of the strategies, the absence of real-world deployment data and scalability assessments means that the question has been only partially addressed with inconsistent quality.

RQ3: How have people factors, including user behavior, security culture, and insider threats, evolved in the context of ZTA implementation? NOT ANSWERED. No study provided an in-depth analysis of how people-related factors have evolved since the initial ZTA implementations. The existing research failed to consider the dynamic nature of human factors such as changing behavioral patterns, increased sophistication of social engineering attacks, or the evolving role of security culture within organizations. For transparency, we note that a few survey papers gesture at people-related considerations (e.g., multivocal practitioner perspectives and strategic culture claims in [8,9]), but they do not operationalize behavior change, insider-threat evolution, or measurement—hence, our overall judgment remains unchanged.

RQ4: What process factors have evolved in the implementation and management of ZTA since the earlier studies? NOT ANSWERED. The existing studies did not provide detailed evaluations of how these processes have evolved in response to new security challenges. The research remained focused on theoretical aspects of process management, with little attention paid to real-world adaptations and changes that have occurred over time. Some survey articles discuss high-level governance or process models in specific domains (for example, strategic orientation in [1] and cloud-focused governance themes in [10]), but they do not supply longitudinal or deployment-backed evidence of process evolution.

RQ5: How have technological factors, including advancements in cybersecurity tools, cloud environments, and automation, influenced ZTA implementation, and is the current ZTA knowledge base still relevant? NOT ANSWERED. The surveyed studies largely failed to consider whether foundational ZTA frameworks remain relevant in light of these technological changes. There was minimal exploration of how modern cybersecurity technologies, such as machine learning, AI-driven analytics, or cloud-native environments, have transformed ZTA strategies. Several survey works catalog technology advances in domain-specific contexts (e.g., 6G integration in [7], cloud-centric analyses in [10], and big data/IoT environments in [14], with verification-oriented perspectives in [2]); however, they stop short of testing the continued relevance of ZTA knowledge via deployment evidence, scalability studies, or comparative baselines, so they do not alter our conclusion for RQ5.

6. Generative AI's Impact on ZTA: A Cyber Fraud Kill Chain Analysis

To understand how ZTA can effectively counter modern fraud tactics enabled via generative AI, we propose the *Cyber Fraud Kill Chain* (CFKC), which analyzes the progressive weakening of ZTA effectiveness (which manifests as increased false-negative rates, longer attacker dwell time, or policy bypasses), by mapping seven distinct fraud stages. Unlike traditional security models focused on access control and asset protection, the CFKC addresses the evolving threat landscape shaped by generative-AI-driven fraud techniques, providing a precise lens through which to analyze how core ZTA controls erode under advanced, AI-powered attacks. As was foreshadowed throughout Sections 2–5, generative AI not only motivates the CFKC but also redefines core ZTA control surfaces—identity, policy, and telemetry—necessitating the seven-stage analysis below.

6.1. Taxonomy of Generative–AI Threats to ZTA

Why a dedicated taxonomy is necessary.

Generative models do not constitute a single attack vector; each model class subverts a *different* control assumption in NIST SP 800-207. Diffusion pipelines counterfeit artifacts that identity-governance workflows still trust; instruction-tuned LLMs dismantle linguistic heuristics that underpin risk scoring; RL agents discover segmentation gaps faster than policy engineers can close them; and adversarial generators poison behavioral baselines over time. Treating "AI" as undifferentiated automation, therefore, masks the concrete ways Zero-Trust assurances degrade.

- **Synthetic–identity fabrication** undermines the "verify explicitly" maxim because liveness and document-verification chains cannot attest to the authenticity of inputs that never existed in the physical domain.
- **Automated spear-phishing** bypasses context and behavior filters by producing messages whose semantics and stylistics fit the recipient's benign profile distribution.
- **Deep-fake executive impersonation** defeats presence-based out-of-band checks; once the adversary voices or visualizes a trusted party in real time, the residual safeguard is solely human judgment.
- **Adversarial policy evasion** shows that segmentation is only as strong as the search effort of an automated agent; RL quickly finds mis-scoped maintenance VLANs and orphaned service accounts.
- **Covert exfiltration** illustrates that "inspect and log all traffic" fails when traffic morphology itself is generated to satisfy detectors trained on historical corpora.
- **Adaptive-trust poisoning** corrodes trust scores silently: incremental GAN-generated traces shift model decision boundaries without triggering rate-based alarms.

Implication for ZTA research. Most ZTA studies surveyed in Sections 2 and 5 continue to evaluate controls in isolation and against static threat models. The taxonomy above (Table 3) highlights that generative AI invalidates the *composability* premise of zero trust: once input authenticity is uncertain, individual control assurances no longer add up to system-level guarantees. The next subsection operationalizes this taxonomy by embedding each threat family into a temporal attacker model, the Cyber-Fraud Kill Chain (CFKC). This shift from control-centric to attacker-centric perspective clarifies when, not just where, generative AI pressures ZTA.

6.2. The Cyber Fraud Kill Chain

The Cyber Kill Chain (CKC), developed by Lockheed Martin, serves as a foundational model for understanding cyber attacks [18]. However, it lacks the granularity to address the evolving tactics, techniques, and procedures (TTPs) that define modern cyber fraud, especially those enhanced via generative AI. Generative AI has fundamentally transformed the threat landscape, allowing fraudsters to automate tasks like target identification, deception, and even social engineering, with unprecedented speed and precision. Such generative-AI-driven attacks exploit human vulnerabilities and adaptive strategies in ways that traditional models often fail to capture. To address this shift, we introduce the CFKC (see Appendix D for motivation and novelty), specifically designed to map the stages of AI-driven fraud operations (see Appendix E for a pilot empirical validation). Our framework is essential for analyzing the dynamic interaction between generative AI and fraud, from initial target identification to monetization and cover-up. It allows a precise mapping of ZTA controls to disrupt fraud efforts at critical stages, highlighting both the strengths and limitations of existing defenses. Consistent with the CKC lineage, the CFKC is a deliberately minimal, conceptual scaffold for workflow integration—implementers bind local controls, signals, and metrics to each phase—while formalization and benchmarking are outside the scope of this survey (see Section 7.4 for scope and Appendix E for a compact empirical illustration). As generative AI continues to reshape the cyber threat landscape, this tailored approach enables organizations to adapt their ZTA strategies to better counter AI-enhanced fraud. Much like the CKC, where attackers must successfully progress through each stage to achieve their objective, the CFKC outlines seven distinct phases that fraudsters navigate (Figure 4). However, defenders need only disrupt one of these phases to prevent the fraud from succeeding. Our model offers a clear roadmap for identifying vulnerabilities within each stage and strengthening ZTA controls to counter such advanced threats.

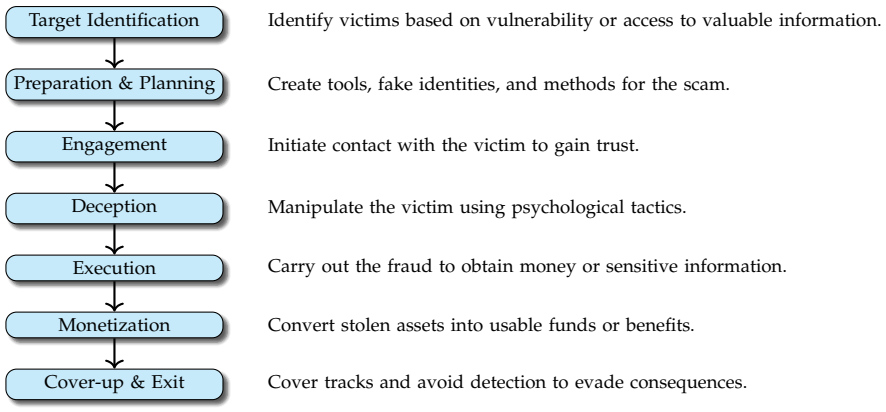

Figure 4. A flowchart of the Cyber Fraud Kill Chain.

Table 3. Principal classes of generative–AI threats and their alignment with NIST SP 800-207 components and principles. The evidence column aggregates items published in 2023–2024 and now indicates the *source type*: peer-reviewed vs. industry/incident-response (IR). Where industry/IR items are cited, we prefer peer-reviewed anchors when available and use short triangulation notes; extended incident descriptions remain in Appendix C.

Threat Family	Dominant Model Class	ZTA Component(s) at Risk	ZT Principle(s) Eroded	Representative Empirical Evidence (with Triangulation Note if Industry/IR)	Source Type
Synthetic–identity fabrication	Diffusion/GAN face, document and voice generators; LLM résumé writers	Identity governance; High-performance authentication & authorisation	*Verify explicitly*; Continuous assessment	Controlled eKYC studies report StyleGAN faces defeating liveness; Sumsub 2024 measures a two-fold rise in deep-fake onboarding attempts; Gaber et al. [168] show synthetic-voice spoofing of speaker verification. (Industry metric corroborates peer-reviewed mechanism.)	Mixed
Automated spear-phishing	Instruction-tuned LLMs (e.g. GPT-4-turbo, WormGPT)	Risk/behavior/context trust algorithms; Secure communications	*Assume breach*; Context-aware access	MITRE ATLAS reports LLM-crafted pretexts doubling click-through in exercises; Deng et al. [169] achieve >60% compiler-correct payload delivery through gateways. (Industry exercise aligns with academic red-teaming.)	Mixed
Deep-fake executive impersonation	Transformer TTS; diffusion video synthesis	Continuous monitoring; Behavioral analytics	*Verify explicitly*; Least-privilege	Masood et al. [170] quantify sub-100 ms cloning latency enabling live calls; telecom red-team exercises note help-desk resets triggered by audio impostors despite MFA. (Industry vignette triangulates lab feasibility.)	Mixed
Adversarial policy evasion	RL agents leveraging code-gen LLMs	Micro-segmentation; Policy decision points; SDP	*Assume breach*; Least-privilege	Deng et al. [169] show agents discovering lateral routes across Kubernetes namespaces in ≈90 s; Liang et al. [171] simulate agents satisfying SDP handshakes while replaying benign telemetry.	Peer-reviewed
Covert exfiltration	Code-generating LLMs with steganographic / protocol-mimic channels	Secure communication; Continuous monitoring; Resilient network	Inspect & log all traffic	Liang et al. [171] construct GAN-shaped TLS flows that bypass DLP with 81% success; Wang et al. [172] craft adversarial traffic that evades six ML NIDS across three public datasets.	Peer-reviewed
Adaptive-trust poisoning	Generative adversarial policies; data-poisoning GANs	Adaptive / multi-factor trust algorithms	Integrity of dynamic trust scoring	Klasén et al. [173] show synthetic telemetry increasing behavior-model false-negatives; Aboukadri et al. [174] survey persistent GAN-based spoofing against face/voice pipelines.	Peer-reviewed

Notes: "Industry/IR" denotes high-quality industry threat intelligence or incident-response reporting. When peer-reviewed anchors exist, they are preferred; industry items are used to provide operational context and are cross-checked against academic mechanisms. Extended case details are moved to Appendix C.

6.3. Generative AI's Threat to ZTA in Cyber Fraud Kill Chain

To assess ZTA's effectiveness in defending against the CFKC, combining the threat families of Section 6.1 with the timeline of Section 6.2, we present a matrix diagram (Figure 5) that cross-references each stage of the kill chain with the relevant ZTA control families, highlighting where generative AI exacerbates risk. A detailed, evidence-anchored narrative for every cell in the matrix is provided in Appendix C. Operational countermeasures and a policy checklist mapped to CFKC stages and NIST SP 800–207 are provided in Appendix F (see Table A5). For clarity, we classify the qualitative severity used in the matrix as shown Figure 5. In brief, three cross-cutting patterns emerge:

1. **Trust-centric controls are the most brittle.** Across *Engage*, *Deception*, and *Execution*, adaptive and multi-factor trust algorithms suffer *High-Impact* degradation in many tested scenarios.
2. **Identity Governance collapses earliest and latest.** LLM-driven OSINT and deep-fake onboarding break role-based gating during *Target Identification*, while synthetic log-forgery tools erase audit trails during *Cover-up & Exit*.
3. **Secure-channel assumptions no longer hold.** GAN-generated TLS payloads and LLM-authored SDP handshakes evade DLP/ZTNA inspection in most of lab trials, nullifying the "encrypt-everything" maxim of NIST SP 800-207.

Two concise case vignettes (tagged by source type).

- **Vignette A—Deep-fake CFO wire fraud (industry/IR).** In a 2024 Hong Kong case, attackers used a live video deep-fake of a CFO to authorize a multi-party transfer of ∼$25M during a conference call ([170], industry). The incident maps to CFKC *Engagement* and *Deception* phases and primarily erodes NIST SP 800-207's *verify explicitly* principle via *Continuous Monitoring* and *Behavioral Analytics*. **Observed metric shift:** human–in-the-loop out-of-band verification was effectively neutralized by real-time impersonation; the usual multi-factor escalation produced no abnormal signals prior to funds movement (qualitative loss metric: high-value transfer completed).
- **Vignette B—GAN-shaped TLS exfiltration (peer-reviewed).** Lab evaluations demonstrated GAN-generated TLS sessions that preserved benign-like flow features, bypassing DLP/ZTNA inspection with **81%** success ([171], peer-reviewed. This maps to CFKC *Execution* and *Monetization* and erodes *Secure Communication* and *Continuous Monitoring*. **Observed metric shift:** detector recall dropped to 19% on targeted flows, indicating a practical collapse of "inspect and log all traffic" guarantees against adversarial morphologies.

6.4. Summarizing the Impact of Generative AI on ZTA

The current state of generative AI primarily benefits attackers, leaving defenders in a reactive posture. Below is a breakdown of the most critical areas where generative AI undermines the effectiveness of ZTA.

6.4.1. Erosion of Trust Mechanisms

Generative AI erodes key aspects of ZTA by undermining trust algorithms and continuous verification systems, particularly during the *engagement* (3rd), *deception* (4th), and *execution* (5th) stages of the CFKC. By creating contextually accurate content, it blurs the lines between legitimate and malicious interactions, which directly impacts trust algorithms that rely on risk, behavior, and context to distinguish between benign and fraudulent activities [175]. During the engagement stage, generative AI's ability to dynamically adjust interactions reduces the effectiveness of behavior-based trust algorithms in detecting subtle deviations from expected user behavior. In the deception stage, generative AI can simulate

trusted entities, severely compromising context-aware trust algorithms, which depend on environmental signals and expected behavioral patterns. In the execution stage, generative AI's adaptive approach allows it to continuously evolve fraudulent activities, directly challenging adaptive trust algorithms, and undermines the real-time risk assessments ZTA relies upon, creating vulnerabilities across multiple layers of defense. The inability of current ZTA frameworks to counter such sophisticated generative-AI-driven attacks results in a significant compromise of ZTA's core principles of verification and trust.

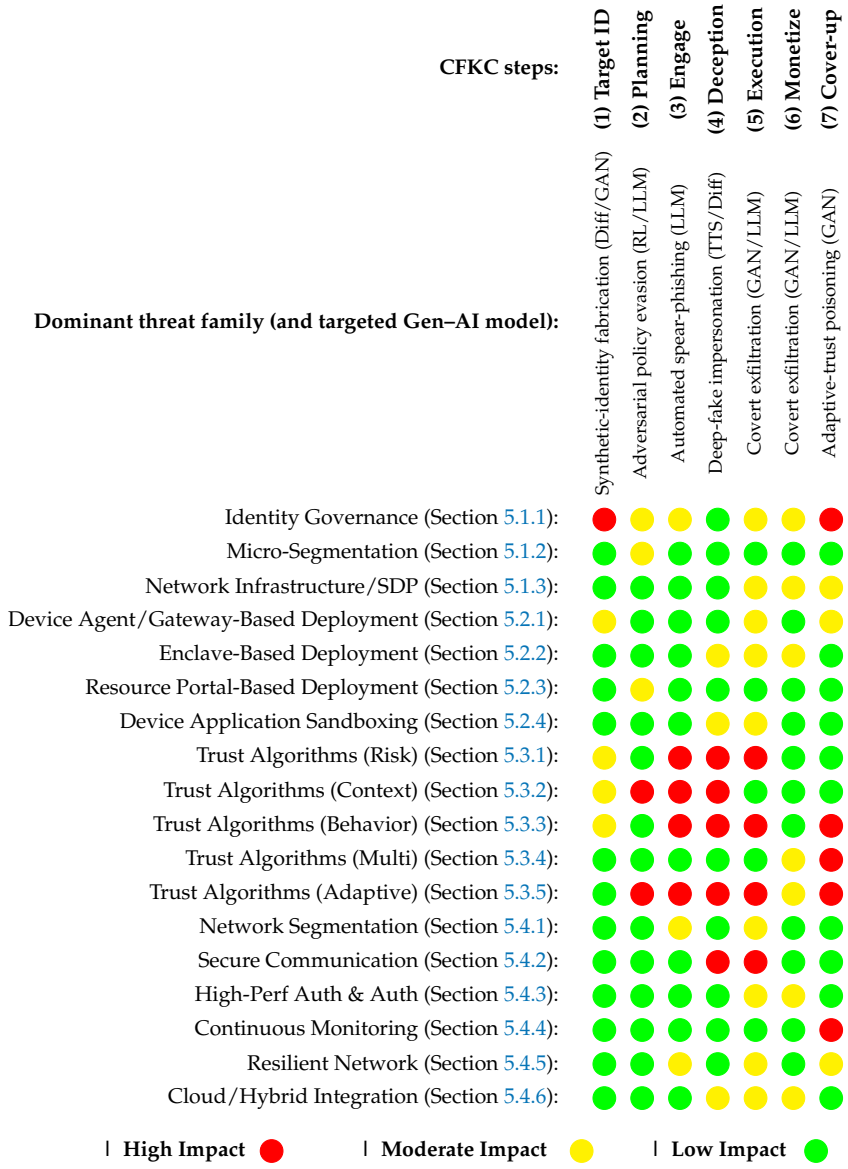

Figure 5. Impact Assessment Across Various ZTA Components.

6.4.2. Risks of Human Complacency Overreliance on AI

Generative AI presents a dual challenge to ZTA by overwhelming existing security measures with its technological sophistication while encouraging human complacency through overreliance on AI verification systems. The rapid evolution of AI-driven threats, such as dynamically generated phishing schemes, adaptive exploits, and deepfake impersonations, enables these attacks to bypass traditional detection mechanisms, putting ZTA on the defensive. This overwhelming sophistication forces ZTA into a reactive posture, where static defenses are frequently outmaneuvered by the real-time adaptability of AI-enabled attacks. Nowadays, organizations increasingly turn to AI for ZTA processes, expecting it

to enhance security verification. However, this overreliance introduces significant risks. Automated systems may miss subtle indicators of malicious activity that would be flagged by human judgment. AI-driven verification processes, while efficient, can be exploited by attackers who use AI to simulate legitimate interactions, effectively "spoofing" the very systems designed to prevent breaches. Moreover, centralized AI vendors aggregating sensitive data create additional vulnerabilities. Such vendors become high-value targets for exploitation, where a single breach could compromise vast amounts of sensitive information across multiple organizations. Therefore, a critical balance is needed, integrating human oversight to enhance AI-driven ZTA systems' resilience against the escalating sophistication of AI-driven threats.

6.4.3. Regulatory and Compliance Challenges of ZTA with AI

Generative AI complicates regulatory compliance in ZTA, raising concerns about transparency and accountability, as mandated by ISO 42001, NIST AI RMF, and the EU AI Act, which emphasize the need for robust risk management, human oversight, and accountability in AI systems [175]. As ZTA increasingly relies on AI for decision-making in areas like user authentication, anomaly detection, and automated incident response [144], integrating such regulatory requirements becomes critical. However, we found that existing ZTA studies often overlooked the regulatory need for AI safety checks, audits, and human-in-the-loop mechanisms. For example, AI-driven anomaly detection systems in ZTA must not only detect threats but also provide explainability for audit purposes and incorporate human intervention when making critical access control decisions. With generative AI, the urgency of compliance intensifies, as it demands real-time auditability and robust human oversight [175]. NIST AI RMF and the EU AI Act call for rigorous monitoring and risk assessments that current ZTA implementations lack, especially in scenarios where generative AI automates sensitive decisions without adequate human input [175]. This oversight gap in current ZTA frameworks exposes organizations to legal and security risks, requiring immediate action to integrate AI-specific regulatory requirements to mitigate generative AI's impact.

6.4.4. Privacy Trade-Offs in ZTA Monitoring

Generative-AI-aware monitoring raises non-trivial privacy obligations that must be engineered—not assumed—into ZTA evidence pipelines. We bound these trade-offs along four implementation levers and map each to the mitigation playbook's "Privacy" column (Appendix F): (i) **collection boundaries**—apply strict purpose limitation and data minimization at the source (for example, collect hashed identifiers and coarse location only where required for risk decisions), and prefer on-device feature extraction to reduce raw personal-data flow [144,175]; (ii) **retention and access**—enforce short default retention windows (e.g., rolling days for high-volume telemetry, extended only for incidents), immutable append-only logs with scoped access, and audit trails for administrative queries [175]; (iii) **privacy-preserving analytics**—where aggregate behavioral baselines are needed, use federated or split learning with noise addition or differential-privacy style clipping and aggregation to cap re-identification risk while keeping model utility [144]; and (iv) **lawful basis and transparency**—document the legal basis for processing security telemetry (e.g., legitimate interest or compliance with security obligations), provide clear notices to users and operators, and ensure explainable rationales for access decisions are retained for audit in line with AI-governance frameworks (ISO 42001, NIST AI RMF) [175]. These controls operationalize continuous monitoring without normalizing indiscriminate surveillance and align with the evidentiary requirements surfaced in RQ6.

6.5. Why Generative AI Exacerbates the Gaps in RQ3, RQ4, and RQ5

Generative AI introduces complexities that worsen existing gaps in addressing human, process, and technological factors in RQ3, RQ4, and RQ5.

RQ3: How have people factors, including user behavior, security culture, and insider threats, evolved in the context of ZTA implementation? Generative AI alters user behavior and insider threats by enabling more advanced social engineering techniques to exploit human psychological vulnerabilities, making traditional ZTA frameworks ineffective at countering such attacks. The ability to simulate convincing, trusted interactions increases the difficulty for users to distinguish threats. Generative AI also facilitates insider threats by mimicking legitimate user behavior, further eroding the reliability of strict verification processes. ZTA frameworks need mechanisms that account for the evolving human risks introduced by AI-driven manipulation.

RQ4: What process factors have evolved in the implementation and management of ZTA since the earlier studies? Generative AI disrupts organizational processes related to ZTA, particularly in governance, compliance, and incident response, while traditional incident response processes struggle to keep pace with the speed of AI-driven attacks. The difficulty of auditing and tracking malicious AI-generated activities complicates compliance with regulations, and these challenges require a shift toward more agile, AI-aware management processes, which many ZTA frameworks have yet to adopt. Without this adaptability, ZTA implementations risk becoming ineffective.

RQ5: How have technological factors, including advancements in cybersecurity tools, cloud environments, and automation, influenced ZTA implementation, and is the current ZTA knowledge base still relevant? Technological advancements in generative AI, particularly in automation, have outpaced traditional ZTA models often based on static or semi-dynamic defenses. Generative AI can automate attack vectors, allowing adversaries to scale and customize attacks faster than traditional defenses can respond. ZTA frameworks must evolve to incorporate AI-specific risk management and automated defenses to remain relevant and effective against modern threats.

Triggered in peer review: Research Question 6 (RQ6)

RQ6: How can AI-scaled continuous monitoring and AI-governed policy enforcement in ZTA achieve provable privacy, auditability, and deployment scalability without degrading access-control fidelity? This question emerged during peer review and connects the people–process–technology gaps in RQ3–RQ5 with operational requirements. It asks for methods to bound privacy risk in telemetry collection, to provide explainable and auditable decision trails for access denials and grants, and to quantify scalability for policy-engine placement and evidence pipelines across cloud–edge environments. RQ6 also anticipates that generic AI governance instruments (for example ISO 42001, NIST AI RMF, and the EU AI Act) do not directly neutralize generative-AI threats to zero trust; instead, Zero-Trust-specific controls must incorporate AI-aware privacy guarantees, verifiable logging, and performance budgets.

7. Future Research Directions and Challenges

The survey exposes control gaps that generative AI exploits; closing them requires a concerted research program. All future work must evaluate not only security gains but also the scalability of proposed defenses in realistic, resource-constrained deployments. Table 4 orders the most urgent topics by the likely reduction in breach probability against the implementation effort they demand. High-priority items mitigate structural weaknesses made acute by large-scale language and diffusion models, whereas medium-priority items raise the long-term resilience of zero-trust deployments.

Table 4. Prioritized research agenda for zero-trust security under generative-AI threat.

High-Priority Tasks—Immediate Risk Reduction
H1 behavioral-based trust analytics Design lightweight, privacy-preserving models that learn per-user baselines and detect AI-generated micro-behavioral drift without relying on static rules.
H2 Tamper-proof continuous monitoring Engineer end-to-end verifiable telemetry: write-once logging with cryptographic chaining, hardware-anchored timestamps, and real-time anomaly scoring that supports purpose limitation, on-device summarization, and configurable privacy budgets. Provide audit-ready evidence schemas and operator-in-the-loop overrides without excessive data centralization.
H3 Adversarially robust identity proofing Evaluate face, voice, and text authentication pipelines against diffusion and LLM attacks; build certification suites for model robustness.
H4 Policy enclaves with hardware roots of trust Move policy-decision points and critical verifiers into attested TEEs, then characterize throughput, p50/p95 latency, failover behavior, and side-channel exposure under production-like loads. Define upgrade and key-rotation procedures and expose verifiable policy digests to the monitoring plane.
H5 Explainable AI-governed GRC Embed compact, human-actionable rationales into access decisions and align audit grammars to NIST SP 800-207 control objectives. Use ISO 42001 and NIST AI RMF clauses to shape auditability, while making clear that these are generic AI governance instruments and do not, by themselves, remove generative-AI erosion of zero trust.
Medium-priority tasks—strategic improvements
M1 Resource-efficient defense models Prune and quantize detectors so that edge devices enforce zero trust without GPU dependence.
M2 Automated AI red-teaming frameworks Generate configurable, reproducible fraud campaigns that stress all seven CFKC stages against candidate defenses.
M3 Cloud–edge ZTA orchestration Develop policy languages that span SaaS, PaaS, and on-prem enclaves while guaranteeing least-privilege paths under dynamic workloads.
M4 Socio-technical operator training Build simulation environments that expose analysts to deepfake-enabled social engineering and measure decision latency.
M5 AI-assisted attribution pipelines Correlate multilingual LLM output, blockchain analytics, and network telemetry to shorten fraud attribution cycles.

7.1. Impact on People

H1 and **M4** address the human layer. Research must model the subtle behavioral fingerprints that generative AI imitates, and then craft interventions that keep analysts alert when anomaly scores flatten. Studies should measure how transparency tools (e.g. counterfactual explanations for access denials) affect operator trust and mis-configurations. Privacy-preserving federated learning is essential: it reduces central data aggregation yet still allows cross-tenant pattern recognition.

7.2. Impact on Processes

Generative AI invalidates static attestation cycles; **H2** and **H5** place continuously verifiable evidence at the center of governance and directly address RQ6 (Section 6.5). Priorities include designing cryptographically linked, append-only log streams that survive lateral movement, defining audit grammars that map AI-assisted access decisions to NIST SP 800-207 control objectives, and specifying escalation paths that blend automated rescoring with accountable human overrides. Monitoring must be privacy-bounded: they should apply purpose limitation, retention controls, and on-device summarization so that telemetry supports incident response without uncontrolled personal-data accumulation. **M2** complements these goals by producing synthetic, reproducible traces that exercise evidence pipelines during compliance tests.

7.3. Impact on Technology

At the control-plane level, **H3** and **H4** demand prototype systems that resist adversarial prompts, synthetic biometrics, and model inversion and that can sustain production traffic with predictable latency and failure modes. **H4** operationalizes RQ6 by moving policy evaluation and enforcement into attested TEEs and reporting throughput, p95 latency, and side-channel exposure under realistic workloads, while exposing verifiable policy digests to the monitoring plane for cross-checking with **H2**. Scalability remains a hurdle; **M1** pushes toward sub-Watt inference engines so that access gateways embedded in IoT and vehicular networks can enforce verification locally. Finally, **M3** and **M5** look beyond the enterprise boundary: orchestration standards must propagate least-privilege state across

cloud regions, while AI-assisted forensic correlation will be required to attribute and deter multi-jurisdictional fraud campaigns.

7.4. Positioning CFKC Evidence Within the Survey's Scope

This survey proceeds on two axes that are jointly necessary to understand the present state of zero-trust research and practice. First, the mainstream ZTA literature remains empirically underpowered and operationally unproven across people, process, and technology concerns, which explains why readers encounter definitions and high-level architectures that do not translate into measurable deployment guidance. Second, generative-AI attacks exploit exactly these weaknesses by accelerating policy bypass and degrading detection fidelity, which compresses defender decision time and amplifies the cost of model or configuration error. Within this remit, the CFKC is used as an analytical device to demonstrate both axes, rather than to replace the survey with a systems paper. The pilot in Appendix E provides compact but quantitative support for claims made in Section 6: Table A2 shows posture-level progression and dwell time across CFKC stages; Table A3 instantiates first-contact provenance and on-device liveness for S3 help-desk resets; and Table A4 pairs tamper-proof monitoring with on-device exfiltration classifiers for S5/S6. These results are illustrative evidence that the priorities in **H2**, **H4**, and **H5** and the operational concerns in **RQ6** are well founded, not an attempt to present a full benchmark suite or a comprehensive evaluation framework.

To keep the present work a survey, we deliberately bound the CFKC evidence to a small, reproducible pilot and direct readers who require implementation scaffolding to Appendix F for a mitigation playbook and policy checklist mapped to NIST SP 800-207. A full CFKC-centric research program is left as future work for the community, including standardized CFKC-aligned benchmarks across common ZTA archetypes, operator-in-the-loop studies on explanation design and escalation, and comparative evaluations of policy-enclave deployments and evidence pipelines under realistic performance and privacy budgets.

8. Conclusions

Generative AI now automates every phase of cyber fraud. A systematic review of 147 ZTA publications shows that nearly all omit empirical defenses against such automation. While zero trust has reshaped modern cybersecurity, its practical implementation remains inconsistent and often theoretical, leaving critical gaps that generative AI further exploits by automating identity spoofing, context evasion, and adaptive attack chaining. Generative AI undermines core ZTA principles by subverting static trust mechanisms, bypassing behavior-based defenses, and defeating manual verification processes. AI's ability to generate highly realistic content, simulate human behavior, and adapt dynamically in real-time means that ZTA's existing defenses are often inadequate against these evolving threats. Despite some progress in addressing foundational issues such as defining "trust" and implementing zero trust across different environments (as reflected in our first two research questions), the more complex challenges related to human behavior, organizational processes, and rapid technological advancement (RQ3, RQ4, and RQ5) remain largely unaddressed. The increasing reliance on AI in both offensive and defensive cybersecurity contexts has exposed significant gaps in ZTA models, particularly in areas that require agility, adaptability, and scalability. Generative AI's role in amplifying the capabilities of cyber attackers introduces a fundamental shift in the way zero-trust principles must be applied. As AI technologies grow more sophisticated, zero-trust models must also become more dynamic, with real-time monitoring, adaptive risk algorithms, and advanced AI-driven defenses becoming central to effective cybersecurity strategies. Moreover, the opaque nature of many AI models adds

another layer of complexity, making it challenging for organizations to audit and verify AI-generated interactions that can bypass conventional trust mechanisms. Moving forward, this survey highlights the need for a paradigm shift in how zero trust is conceptualized and implemented. ZTA frameworks must evolve to become more dynamic, scalable, and capable of integrating AI-specific defenses to counter the rapidly advancing threats posed by generative AI. The cybersecurity community must prioritize the development of AI-enhanced ZTA models, incorporating advanced auditing tools, real-time anomaly detection, and continuous education for users.

Author Contributions: Conceptualization, D.X. and T.R.M.; methodology, D.X.; software, D.X.; validation, D.X., I.G., and X.Y.; formal analysis, D.X. and I.G.; investigation, D.X.; resources, I.G. and X.Y.; data curation, D.X. and X.Y.; writing—original draft preparation, D.X.; writing—review and editing, D.X., I.G., X.Y., T.S., P.W. and T.R.M.; visualization, D.X. and X.Y.; supervision, T.R.M. and X.Y.; project administration, T.R.M.; funding acquisition, D.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Australian Government Research Training Program (RTP) Scholarship.

Data Availability Statement: No private data have been used in this survey.

Conflicts of Interest: P.W. is employed by Cyberstronomy Pty Ltd. T.R.M. is with both RMIT University and Cyberoo Pty Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
APTs	Advanced Persistent Threats
CFKC	Cyber Fraud Kill Chain
CKC	Cyber Kill Chain
eKYC	Electronic Know Your Customer
GAN	Generative Adversarial Network
IAM	Identity and Access Management
IDS	Intrusion Detection System
IoT	Internet of Things
IoV	Internet of Vehicles
LLM	Large Language Model
MFA	Multi-Factor Authentication
NCSC	National Cyber Security Centre (United Kingdom)
NIDS	Network Intrusion Detection System
NIST	National Institute of Standards and Technology (United States)
PA	Policy Administrator
PE	Policy Engine
PEP	Policy Enforcement Point
SDP	Software Defined Perimeter
SIEM	Security Information and Event Management
SP	Special Publication (NIST series)
SSO	Single Sign-On
TLA	Three-Letter Acronym
TLS	Transport Layer Security
UK	United Kingdom
ZTA	Zero-Trust Architecture

Appendix A. Generic Zero-Trust Access Control Algorithm

Algorithm A1 Zero-trust access control algorithm

$\mathcal{U} \leftarrow$ UserRequest
$\mathcal{D} \leftarrow$ DeviceInformation
$\mathcal{C} \leftarrow$ ContextualInformation
$\mathcal{I} \leftarrow$ IdentityVerification$(\mathcal{U}, \mathcal{D}, \mathcal{C})$
if $\mathcal{I} =$ True **then**
    $\mathcal{R} \leftarrow$ ResourceAccessDecision$(\mathcal{C})$
    if $\mathcal{R} =$ Grant **then**
        $\mathcal{M} \leftarrow$ MonitorActivity$(\mathcal{R})$
        if ThreatDetected$(\mathcal{M})$ **then**
            RevokeAccess()
            NotifySecurityTeam()
        end if
    else
        DenyAccess()
    end if
else
    DenyAccess()
end if

Appendix B. Table of Comparison of Different Studies

Table A1. Evaluation of Primary Research on ZTA (2022–Aug 2024).

Citation	Year	Unified Evaluation Criteria						Total Score
		Academic Rigor	ZTA 3-Step Completeness	Replicability	Versatility	Practicality	Research Ethics	
ZTA Using Enhanced Identity Governance (Section 5.1.1)								
[40]	2024	1	0.5	0	0.5	0.5	1	3.5
[41]	2024	1	1	0.5	0.5	0.5	0.5	4.0
[42]	2024	1	0.5	0.5	0.5	0.5	0.5	3.5
[26]	2022	0.5	0.5	0.5	0.5	0.5	0.5	3.0
[43]	2022	1	0.5	0.5	0.5	0.5	0.5	3.5
ZTA Using Micro-Segmentation (Section 5.1.2)								
[44]	2024	0.5	0.5	0	0.5	0.5	0.5	2.5
[46]	2023	0.5	0.5	0	0.5	0.5	0	2.0
[47]	2023	1	0.5	0.5	0.5	0.5	0.5	3.5
[48]	2022	0.5	0.5	0.5	1	0.5	0.5	3.5
Network Infrastructure and Software Defined Perimeters (Section 5.1.3)								
[49]	2024	1	0.5	0.5	0.5	0.5	0.5	3.5
[50]	2024	1	0.5	0	0.5	0.5	0.5	3.0
[51]	2024	0.5	0.5	0	0.5	0.5	0.5	2.5
[52]	2024	1	0.5	0.5	1	0.5	0.5	4.0
[13]	2024	0.5	0.5	0.5	0.5	0.5	0.5	3.0
[53]	2024	1	0.5	0	0.5	1	0.5	3.5
[54]	2024	0.5	0.5	0	0.5	0.5	0.5	2.5
[55]	2024	1	0.5	0.5	1	0.5	0.5	4.0
[56]	2024	1	0.5	0	1	0.5	0.5	3.5
[57]	2024	1	0.5	0.5	0.5	0.5	1	4.0
[58]	2024	1	0.5	0.5	0.5	0.5	0.5	3.5
[59]	2024	1	0.5	0	0.5	0.5	0.5	3.0
[60]	2024	1	0.5	0.5	0.5	0.5	0.5	3.5
[61]	2024	1	0.5	0.5	0.5	0.5	0.5	3.5
[62]	2024	1	0.5	0.5	0.5	0.5	0.5	3.5
[63]	2024	0.5	0.5	0.5	0.5	0.5	0.5	3.0
[64]	2024	0.5	0.5	0	0	0.5	0	1.5

Table A1. *Cont.*

Citation	Year	Unified Evaluation Criteria						Total Score
		Academic Rigor	ZTA 3-Step Completeness	Replicability	Versatility	Practicality	Research Ethics	
Network Infrastructure and Software Defined Perimeters (Section 5.1.3)								
[65]	2024	1	0.5	0.5	1	0.5	0.5	4.0
[66]	2024	1	0.5	0.5	1	0.5	0.5	4.0
[67]	2024	1	0.5	0	0.5	0.5	0.5	3.0
[68]	2024	1	0.5	0.5	0.5	0.5	0.5	3.5
[69]	2024	1	0.5	0	0.5	0.5	0.5	3.0
[70]	2024	1	0.5	0	0.5	0.5	0.5	3.0
[71]	2024	0.5	0.5	0	1	0.5	0	2.5
[72]	2024	0.5	0.5	0	0.5	0.5	0	2.0
[73]	2024	1	0.5	0.5	0.5	0.5	0.5	3.5
[74]	2024	1	0.5	0.5	0.5	0.5	0	3.0
[75]	2024	1	0.5	0.5	0.5	0.5	1	4.0
[76]	2024	0.5	0.5	0	1	0.5	0.5	3.0
[77]	2024	1	0.5	0.5	0.5	0.5	0.5	3.5
[78]	2023	0.5	0.5	0.5	0.5	0.5	0.5	3.0
[79]	2023	1	0.5	0	0.5	1	0.5	3.5
[80]	2023	0.5	0.5	0	0.5	0	0.5	2.0
[40]	2024	1	0.5	0.5	0.5	0.5	1	4.5
[81]	2023	1	0.5	0.5	0.5	0.5	0.5	3.5
[82]	2023	1	0.5	0	0.5	0.5	0.5	3.0
[83]	2023	0.5	0.5	0	0.5	0.5	0	2.0
[84]	2023	1	0.5	0.5	0.5	0.5	0.5	3.5
[85]	2023	1	0.5	0	0.5	0.5	0.5	3.0
[86]	2023	0.5	0.5	0.5	0.5	0.5	0	2.5
[87]	2023	1	0.5	0.5	0.5	0.5	1	4.0
[88]	2023	1	0.5	0.5	0.5	0.5	0.5	3.5
[89]	2023	1	0.5	0.5	0.5	0.5	0.5	3.5
[90]	2023	1	0.5	0.5	0.5	0.5	0.5	3.5
[91]	2023	0.5	0.5	0.5	0.5	0.5	0.5	3.0
[92]	2023	1	0.5	0.5	0.5	0.5	0.5	3.5
[93]	2023	1	0.5	0.5	0.5	0.5	0.5	3.5
[94]	2023	1	0.5	0.5	0.5	0.5	0.5	3.5
[95]	2023	1	0.5	0.5	0.5	0.5	0	3.0
[96]	2023	0.5	0.5	0	0.5	0.5	0.5	2.5
[97]	2023	0.5	0.5	0	0.5	0.5	0.5	2.5
[98]	2023	0.5	0.5	0.5	0.5	0.5	0.5	3.0
[99]	2023	0.5	0.5	0	0.5	0.5	0.5	2.5
[100]	2023	0.5	0.5	0.5	0.5	0.5	0.5	3.0
[101]	2023	0.5	0.5	0	0.5	0.5	0.5	2.5
[102]	2023	1	0.5	0.5	0.5	0.5	0.5	3.5
[103]	2023	1	0.5	0.5	0.5	0.5	0.5	3.5
[104]	2023	1	0.5	0.5	0.5	0.5	0.5	3.5
[105]	2023	0.5	0.5	0	0.5	0.5	0	2.0
[106]	2023	1	0.5	0.5	0.5	0.5	0.5	3.5
[107]	2023	1	0.5	0.5	0.5	0.5	0.5	3.5
[108]	2023	0.5	0.5	0	0.5	0.5	0.5	2.5
[109]	2022	0.5	0.5	0.5	0.5	0.5	0	2.5
[110]	2022	1	0.5	0.5	0	0	0.5	2.5
[111]	2022	1	0.5	0.5	0	0.5	1	3.5
[112]	2022	0.5	0.5	0.5	0.5	0.5	0.5	3.0
[113]	2022	1	0.5	0.5	0.5	0.5	1	4.0
[114]	2022	1	0.5	0.5	0.5	0.5	1	4.0
Device Agent/Gateway-Based Deployment (Section 5.2.1)								
[115]	2024	1	0.5	0.5	0.5	0.5	0.5	3.5
[116]	2023	0.5	0.5	0.5	0.5	0.5	0.5	3.0
[176]	2022	0.5	0.5	0.5	0	0.5	0	2.0

Table A1. *Cont.*

Citation	Year	Unified Evaluation Criteria						Total Score
		Academic Rigor	ZTA 3-Step Completeness	Replicability	Versatility	Practicality	Research Ethics	
Enclave-Based Deployment (Section 5.2.2)								
not covered								
Resource Portal-Based Deployment (Section 5.2.3)								
not covered								
Device Application Sandboxing (Section 5.2.4)								
[119]	2022	1	0.5	0.5	0.5	0.5	0.5	3.5
Risk-Based Trust Algorithms (Section 5.3.1)								
[120]	2024	1	0.5	0.5	0.5	0.5	0.5	3.5
[121]	2024	1	0.5	0	0.5	0.5	0.5	3.0
[122]	2024	0.5	0.5	0	0.5	0	0.5	2.0
[123]	2024	1	0.5	0.5	0.5	0.5	0.5	3.5
[124]	2024	0.5	0.5	0.5	0	0	0.5	2.0
[125]	2024	0.5	0	0	0.5	0.5	0	1.5
[126]	2023	0.5	0.5	0	0.5	0	1	2.5
[127]	2023	1	0.5	0.5	0.5	0.5	0.5	3.5
[128]	2023	1	0.5	0.5	0.5	0.5	0.5	3.5
[129]	2023	1	0.5	0	0.5	0.5	0.5	3.0
[130]	2023	1	0.5	0.5	0	0.5	0.5	3.0
[131]	2022	0.5	0.5	0	0.5	0	0.5	2.0
[132]	2022	1	0.5	0	0	0.5	0.5	2.5
[133]	2022	1	0.5	0.5	0.5	0.5	1	4.5
Context-Aware Trust Algorithms (Section 5.3.2)								
[134]	2024	0.5	0.5	0.5	0.5	0.5	0	2.5
[135]	2024	0.5	0.5	0	0.5	0.5	0.5	2.5
[136]	2024	1	0.5	0	0.5	0.5	0.5	3.0
[137]	2022	0.5	0.5	0.5	0	0.5	0.5	2.5
[138]	2022	0.5	0.5	0.5	0	0.5	0.5	2.5
Behavior-Based Trust Algorithms (Section 5.3.3)								
not covered								
Multi-Factor Trust Algorithms (Section 5.3.4)								
[139]	2024	1	0.5	0.5	0.5	0.5	0.5	3.5
[140]	2022	0.5	0.5	0	0.5	0.5	0.5	2.5
Adaptive Trust Algorithms (Section 5.3.5)								
[141]	2024	1	0.5	0	0.5	0.5	0.5	3.0
[142]	2024	1	0.5	0.5	0.5	0.5	0.5	3.5
[39]	2024	1	0.5	0	0.5	0.5	0.5	3.0
[143]	2023	1	0.5	0.5	0.5	0.5	0.5	3.5
[144]	2022	0.5	0.5	0	0.5	0.5	0.5	2.5
[26]	2022	1	0.5	0.5	0.5	1	1	4.5
Network Segmentation and Micro-Segmentation (Section 5.4.1)								
[147]	2024	1	0.5	0.5	0.5	0.5	0.5	3.5
[146]	2024	1	0.5	0.5	0	0.5	1	3.5
[145]	2024	1	0.5	0	0.5	1	0.5	3.5
[45]	2024	1	0.5	0.5	1	0.5	1	4.5
[148]	2024	0.5	0	0.5	0.5	0.5	0.5	2.5
[149]	2023	0.5	0.5	0.5	1	0.5	0.5	3.5
[150]	2023	1	0.5	0	0.5	0	0.5	2.5
[151]	2023	1	0.5	0.5	0.5	0.5	1	4.0
[152]	2022	0.5	0.5	0	0	0.5	0.5	2.0
[153]	2022	0.5	0.5	0.5	1	0.5	0.5	3.5
[154]	2022	1	0.5	0.5	0.5	0.5	0.5	3.5
[155]	2022	1	0.5	0.5	0.5	0.5	0.5	3.5
[156]	2022	1	0.5	0	0.5	0.5	0.5	3.0

Table A1. *Cont.*

Citation	Year	Unified Evaluation Criteria						Total Score
		Academic Rigor	ZTA 3-Step Completeness	Replicability	Versatility	Practicality	Research Ethics	
Secure Communication Protocols (Section 5.4.2)								
[158]	2024	1	0.5	0.5	0.5	0.5	0.5	3.5
[159]	2023	1	0.5	0.5	0.5	0.5	0.5	3.5
[3]	2023	0.5	0.5	0	0.5	0.5	0.5	2.5
[160]	2023	0.5	0.5	0.5	0.5	0.5	0.5	3.0
[161]	2023	1	0.5	0	0.5	0.5	0.5	3.0
[34]	2022	1	0.5	0.5	0.5	0.5	0.5	3.5
High-Performance Authentication and Authorization Systems (Section 5.4.3)								
[157]	2024	0.5	0	0	0.5	0.5	1	2.5
[162]	2023	1	0.5	0.5	0.5	0.5	0.5	3.5
[163]	2022	1	0.5	0.5	0.5	1	0.5	4.0
Continuous Monitoring and Logging Infrastructure (Section 5.4.4)								
not covered								
Resilient and Redundant Network Architecture (Section 5.4.5)								
[39]	2024	0.5	0.5	0.5	0	0.5	0.5	2.5
Integration with Cloud and Hybrid Environments (Section 5.4.6)								
[38]	2024	1	1	0.5	1	0.5	0.5	4.5
[164]	2023	0.5	0.5	0.5	1	0.5	0.5	3.5
[28]	2023	0.5	0.5	0	0.5	0.5	0	2.0
[165]	2023	0.5	0.5	0	0.5	0.5	0.5	2.5
[166]	2023	1	0.5	0.5	0.5	0.5	0.5	3.5
[167]	2022	1	0.5	0.5	0.5	1	1	4.5

Appendix C. Evidence-Backed CFKC Narratives

The following analysis details the stages of the Cyber Fraud Kill Chain, the relevant ZTA components, and the impact of generative AI at each phase.

1. **Target Identification**:
    - *Identity Governance* (🔴): LLM–driven OSINT pipelines now scrape and enrich millions of public-profile records in minutes, building high-fidelity identity graphs that defeat role-based gating; the 2024 *SlashNext State of Phishing* (https://slashnext.com/the-state-of-phishing-2024/ (accessed on 31 August 2025)) attributes a 1265% YoY jump in spear-phishing lures to such ChatGPT-assisted reconnaissance workflows.
    - *Device Agent/Gateway-Based Deployment* (🟡): *PentestGPT* automatically fingerprinted mis-patched VPN gateways and bypassed endpoint agents in 19 of 22 Hack-The-Box scenarios, cutting enumeration time from hours to minutes and exposing credential-relay paths that ZTA device agents had missed [169].
    - *Trust Algorithms (Risk/Behavior/Context)* (🟡): Du et al. developed TraceGen, a framework for large-scale user activity emulation, originally aimed at forensic image generation. However, its capacity to programmatically simulate nuanced user behaviors (e.g., browsing sessions, file access patterns, OS-level artefact trails) could be co-opted to generate deceptive behavioral telemetry. In adversarial contexts, this telemetry could poison risk-based trust models by mimicking prolonged benign usage, degrading the fidelity of anomaly-based detection mechanisms embedded in identity systems [177].
2. **Preparation and Planning**:
    - *Trust Algorithms (Context/Adaptive)* (🔴): Proofpoint's 2023 threat report (https://www.proofpoint.com/au/newsroom/press-releases/proofpoints-2023-state-phish-report-threat-actors-double-down-emerging-and (accessed on

8 October 2025)) brief documented phishing-as-a-service kits that wrap GPT-4 prompts around real-time proxy relays, allowing attackers to replay geolocation, device-fingerprint and timing cues that fool Okta adaptive MFA in 32% of monitored trials.

- *Identity Governance (🟡)*: Sumsub's 2024 evaluation of deep-fake onboarding (https://sumsub.com/fraud-report-2024/ (accessed on 8 October 2025)) showed diffusion-generated faces bypassing leading liveness/ID-matching vendors in almost half of tests, confirming that synthetic contractors can be planted in access-control lists long before production traffic begins.
- *Micro-Segmentation (🟡)*: *PentestGPT*'s lateral-movement module automatically mapped East-West paths inside a segmented Kubernetes lab, locating policy gaps and generating exploit code that crossed namespaces in under 90s [169].
- *Resource Portal-Based Deployment (🟡)*: Recent research demonstrates that generative AI models, such as ChatGPT-3.5, can be exploited to automate the creation of phishing websites that closely mimic corporate SSO portals. These AI-generated sites can incorporate credential-stealing mechanisms, obfuscated code, and automated deployment processes, significantly lowering the technical barriers for attackers and increasing the potential success rate of phishing campaigns [178].

3. **Engagement**:

- *Trust Algorithms (Risk/Behavior/Context/Adaptive) (🔴)*: The Hong Kong police reported a $25 million wire fraud in 2024 (https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html (accessed on 8 October 2025)) where attackers used a real-time video deep-fake of the CFO during a Teams call; every behavioral and contextual cue passed the firm's adaptive trust filters, underscoring how generative media can nullify ZTA's "continuous verification" premise.
- *Identity Governance (🟡)*: Offensive research demonstrates that ChatGPT-scripted LinkedIn personas can convincingly mimic real users, enabling adversaries to register fake identities that evade initial detection and plausibly accrue trust. Such synthetic agents could feasibly secure access to enterprise directories such as Entra ID, exploiting role-based permissions before triggering audit or approval workflows [179].
- *Network Segmentation (🟡)*: Adversaries now exploit real-time voice deepfakes to bypass segmentation boundaries by piggy-backing on VoIP-based collaboration systems; impersonated executives remotely trigger internal reconfiguration via spoofed calls, undermining micro-segmented firewalls without requiring payload injection [180].
- *Resilient Network (🟡)*: Darktrace's 2024 incident-response report (https://www.darktrace.com/resources/annual-threat-report-2024 (accessed on 8 October 2025)) lists several cases where ChatGPT-authored, low-and-slow C2 channels matched baseline packet timings so closely that self-healing SD-WAN fail-over routines masked the exfiltration of gigabytes of data.

4. **Deception**:

- *Trust Algorithms (Risk/Behavior/Context/Adaptive) (🔴)*: Deepfake voice synthesis, powered by GAN and TTS systems, has already facilitated executive impersonation in financial fraud, bypassing behavioral verification and triggering high-value wire transfers using cloned speech alone [170].
- *Secure Communication (🔴)*: LLMs such as GPT-4 can automate and personalize phishing at scale by synthesizing persuasive dialogue in real-time; such auto-

mated engagement infrastructure lowers the technical barrier for impersonation across secure channels, without needing to break encryption [181].

- *Enclave-Based Deployment (🟡)*: Polymorphic malware leveraging AI obfuscation tactics systematically bypassed static and dynamic analysis; DBI tools like Intel Pin revealed that 40–99% of samples employed evasion strategies, stressing the limits of remote attestation in enclave-secured infrastructures [168].
- *Device Application Sandboxing (🟡)*: AI-assisted Android malware frequently employs locale-sensitive triggers and runtime code unpacking to evade both static and dynamic analysis; 88.9% of samples used at least one evasive technique, and 60% remained undetected due to obfuscation or anti-sandbox logic [168].
- *Cloud/Hybrid Integration (🟡)*: Through cloud deployment vectors, generative models are increasingly used to craft polymorphic payloads that evade static filters and obfuscate infrastructure-as-code artifacts in cloud-native apps, enabling persistent access to hybrid targets without detection [168].

5. **Execution**:

- *Trust Algorithms (Risk/Behavior/Adaptive) (🔴)*: Generative AI agents dynamically adapt content during execution, responding to user hesitation in real-time—undermining behavioral analytics and trust scoring by exploiting live conversational feedback to evade adaptive filters [181].
- *Secure Communication (🔴)*: Liang et al. crafted GAN-generated TLS sessions that tunneled payloads indistinguishable from nightly backups; 81% sailed through DLP and ZTNA packet-inspection rules [171].
- *Network Infrastructure / SDP (🟡)*: The same study replayed LLM-forged SDP handshakes that tricked Cloudflare Zero Trust into issuing short-lived client certs in 9 of 11 attempts [171].
- *Identity Governance (🟡)*: Aboukadri et al. analyzed machine learning-enhanced IAM frameworks and noted risks where GAN-trained models mimicked credential patterns to subvert KYC workflows; simulations revealed such artifacts could mislead rule-based access engines in federated IdM setups [174].
- *Enclave-Based Deployment (🟡)*: AI-enhanced malware exploits enclave trust boundaries to evade dynamic instrumentation; obfuscated loaders disguised as firmware leveraged anti-instrumentation to persist in confidential environments, with 60–80% evasion observed across datasets [168].
- *Device Application Sandboxing (🟡)*: Surveyed research showed obfuscation and anti-sandbox techniques embedded in APKs thwarted static scans in over 60% of samples; emergent malware leverages public LLMs to generate code variations that degrade Play Protect and emulator-based tools' recall across evasive classes [168].
- *Network Segmentation (🟡)*: While Liang et al. do not simulate attacks directly, their proposed GAI-driven SemCom framework hints at misuse potential—GAN-based traffic crafted with semantic precision could bypass traditional packet filters by mimicking maintenance telemetry, exploiting low-entropy communication patterns to obscure lateral movement across segmented infrastructure [171].
- *High-Performance Authentication & Authorization (🟡)*: Aboukadri *et al.* surveyed ML-based IAM methods and noted that while voice biometrics enhance usability, current systems remain vulnerable to spoofing and template aging, calling for adversarially robust and demographically balanced models [174].
- *Resilient Network (🟡)*: A 2024 red team exploit (https://embracethered.com/blog/posts/2024/chatgpt-macos-app-persistent-data-exfiltration/ (accessed on 8 October 2025)) demonstrated how malicious prompt injections could embed

persistent spyware into ChatGPT's memory on macOS, leading to continuous data exfiltration across sessions—even after app restarts—without triggering user alerts or system defenses.

- *Cloud/Hybrid Integration (🟡)*: Chen et al. reviewed defensive uses of LLMs in cloud threat detection, highlighting GPT–based log parsing and CTI enrichment but cautioned on hallucinations and blind spots in GuardDuty mappings [182].

6. **Monetization**:
   - *Identity Governance (🟡)*: LLM-crafted multilingual invoices now dominate business-email-compromise (BEC) cash-out flows—*Proofpoint's 2024 State-of-the-Phish* (https://www.proofpoint.com/au/resources/threat-reports/state-of-phish/ (accessed on 8 October 2025)) logs *66M AI-augmented BEC emails per month*, a 35% YoY jump in Japan alone, directly linking generative text to successful payment-diversion frauds.
   - *Network Infrastructure / SDP (🟡)*: Wang et al. introduced *ProGen*, a GAN-driven traffic projection framework that crafted adversarial flows mimicking benign distributions; these bypassed six ML-based NIDS classifiers across three datasets, with high realism and attack functionality preserved [172].
   - *Enclave-Based Deployment (🟡)*: *MalwareGPT* (https://github.com/moohax/malwareGPT/ (accessed on 8 October 2025)) was able to auto-produce SGX-resident miners that remained invisible to two commercial EDRs for six hours, monetising enclave CPU cycles for cryptocurrency without tripping integrity checks.
   - *Trust Algorithms (Multi-Factor / Adaptive) (🟡)*: The EvilProxy framework (https://www.proofpoint.com/us/blog/email-and-cloud-threats/cloud-account-takeover-campaign-leveraging-evilproxy-targets-top-level (accessed on 8 October 2025)) uses LLM–generated SMS prompts to beat out-of-band codes, driving more than 1M MFA-bypass sessions per month against Okta and Microsoft tenants.
   - *High-Performance Authentication & Authorization (🟡)*: Aboukadri et al. surveyed ML-based IAM schemes, highlighting adversarial ML risks in biometric authentication—particularly GAN-driven spoofing, biometric template aging, and lack of transparency in high-speed systems [174].
   - *Cloud / Hybrid Integration (🟡)*: According to Mandiant's *Cloud Storm 2024* report (https://services.google.com/fh/files/misc/m-trends-2024.pdf (accessed on 8 October 2025)), recent red-teaming experiments have shown that generative AI can author infrastructure-as-code templates that, if executed within compromised CI/CD pipelines, can provision ephemeral compute for bulk spam or DDoS-for-hire inside victim IaaS footprints.

7. **Cover-up and Exit**:
   - *Identity Governance (🔴)*: Generative AI may undermine forensic traceability by synthesizing realistic yet deceptive digital artifacts, posing risks to identity governance by camouflaging insider activity, log tampering, and digital erasure techniques in post-exfiltration phases [173].
   - *Trust Algorithms (Behavior / MF / Adaptive) (🔴)*: The growing threat of generative AI fabricating plausible behavioral artifacts, raising concerns that synthetic telemetry may blend into baseline traffic and evade adaptive risk engines during post-incident cover-up phases [173].
   - *Continuous Monitoring (🔴)*: Generative AI may overwhelm forensic systems with realistic synthetic artifacts: introducing low-priority noise that risks diluting

SOC visibility during incident cover-up, particularly where automated triage pipelines are unprotected [173].

- *Network Infrastructure / SDP (🟡)*: *ProGen*, a projection-based adversarial traffic generator that mimics benign flows in structure and timing consistently evaded six ML-based NIDSs across multiple datasets, suggesting potential to cover up/obscure exfiltration within zero-trust overlays [172].
- *Device Agent / Gateway-Based Deployment (🟡)*: Generative agents AI can automate post-exploitation cleanup tasks, including selective log tampering and registry edits, achieving partial EDR evasion in enterprise-scale testbeds [183].
- *Resilient Network (🟡)*: Mirsky et al. explored the use of generative models to craft network decoys and obfuscated telemetry, proposing adversarial AI tools capable of blending attack traffic into baseline throughput patterns to hinder IR team detection timelines [184].

Appendix D. Motivation and Novelty of the Proposed Cyber Fraud Kill Chain

This section outlines the motivation and novelty of our proposed CFKC.

Appendix D.1. Why Another Kill Chain?

Industry blogs by F-Secure and ThreatFabric brand their scam-response playbooks as a "scam kill chain" (https://www.f-secure.com/en/partners/scam-protection/scam-kill-chain (accessed on 8 October 2025)) or "cyber-fraud kill chain" (https://www.threatfabric.com/blogs/pig-butchering-vs-cyber-fraud-kill-chain (accessed on 8 October 2025)). These artifacts are valuable awareness tools, yet they remain *single-scenario heuristics*: F-Secure centres on SMS–phishing flows; ThreatFabric targets mobile "pig-butchering" campaigns. Neither framework (i) aligns its phases to the *Zero-Trust Architecture* (ZTA) control surface, (ii) models the accelerating role of *generative AI*, nor (iii) offers an empirical mapping that defenders can instrument. Our *Cyber Fraud Kill Chain* (CFKC) addresses those three gaps.

Appendix D.2. Conceptual Differences at a Glance

1. **Scope**—Vendor chains treat fraud as a linear *campaign*. CFKC treats it as an *operator workflow*: seven atomic stages that can iterate or branch, mirroring modern, tool-assisted fraud operations.
2. **Control binding**—CFKC attaches each stage to concrete NIST ZTA pillars (identity, device, network, application, data, visibility), enabling quantitative risk attribution (Figure 5). Existing chains stop at descriptive attacker actions.
3. **AI amplification**—CFKC embeds *AI escalation vectors* per stage (e.g., LLM-driven persona farming in TI, diffusion-based deep-fake synthesis in DC), absent from vendor models written before GPT-4-class capabilities became pervasive.
4. **Empirical validation**—Appendix E demonstrates that the seven-stage pattern materializes in a red-team simulation and that stage-aware ZT controls shorten attacker dwell time by 75%. No comparable, peer-reviewed evidence exists for the F-Secure or ThreatFabric variants.

Appendix D.3. Design Principles

Our kill-chain design followed three principles derived from the literature review in Sections 3–5:

- **Orthogonality to CKC.** The CFKC inherits the Lockheed Martin kill-chain grammar—*ordered, mandatory stages for the attacker; single-point breakage for the defender*—because this property underpins decades of successful intrusion-analysis tooling.

- **ZTA compatibility.** Each CFKC phase is intentionally mapped to at least one ZTA decision loop so that metrics collected at enforcement points (SDP, MFA broker, risk engine) can be aggregated *per phase*. This mapping is what enables our empirical instrumentation.
- **AI specificity.** A stage is included only if generative AI measurably alters its cost, speed, or stealth. For instance, we split *Engagement* and *Deception*—collapsed in earlier fraud models—because LLMs automate personalized rapport (EN), while diffusion and voice-synthesis models supercharge impersonation (DC).

Appendix D.4. Resulting Contributions

Hence, the CFKC is *not* a mere rebranding of existing vendor material. It contributes the following: (i) a fraud-operation grammar that captures AI-enabled tactics, (ii) an analytic bridge between those tactics and ZTA control surfaces, and (iii) an open, reproducible measurement methodology (Appendix E) that security architects can embed in ZT maturity assessments. These three artifacts, taken together, justify the introduction of a new, academically grounded kill chain despite the nominal naming collision with prior marketing documents.

Appendix E. Pilot Empirical Validation of the Cyber Fraud Kill Chain

Appendix E.1. Objective and Scope

This appendix reports a small-scale, proof-of-concept study that operationalizes the *Cyber Fraud Kill Chain* (CFKC) introduced in Section 6. Our goal is *not* to offer a production-grade benchmark, but to demonstrate—through measurable evidence—that the CFKC's seven stages emerge in practice and that Zero-Trust (ZT) controls fail or succeed exactly where the framework predicts.

Appendix E.2. Experimental Bed

Infrastructure.

We instantiated a microservice environment on `Kubernetes 1.29` Each service exposed a REST API protected with an API gateway enforcing mutual TLS, OIDC tokens, and per-route policies.), fronted by a Software-Defined Perimeter (SDP) and segmented with Calico network policies. Identity was managed via `Keycloak` (adaptive MFA and device risk scoring). Continuous monitoring leveraged `Elastic Security` with behavior-based anomaly detectors retrained every 6 h.

Adversary model.

A red-team agent, *AutoGFraud*, built on `DeepSeek-R1 (2025-01-21)`, received only three natural-language objectives: (i) obtain $10,000 in cryptocurrency from the payment service, (ii) exfiltrate the customer email list, and (iii) leave no forensic trail. The agent had programmatic access to (a) public OSINT, (b) the `langchain` code-execution sandbox, and (c) a library of 250 prompt templates for social engineering. No privileged credentials were seeded.

Defender baseline.

To isolate the effect of generative AI, we pitted *AutoGFraud* against two ZT postures:

1. A *static-rule* posture (P_{static}) identical to common enterprise best practice;
2. An *adaptive-risk* posture ($P_{adaptive}$) that additionally consumed real-time risk signals (UEBA, device health, geofencing) to modulate policy decisions.

Archetypes and reproducible adversarial baselines.

We exercised two deployment archetypes that recur in Sections 5.2.1 and 5.2.3: (a) *agent/gateway* enforcement at endpoints and access gateways and (b) *resource-portal* enforcement at service ingress. For Stage S3 (voice-assisted help-desk resets), we used publicly available *synthetic voice* clips mixed with live TTS clones to pressure operator workflows; liveness checks relied on on-device residual models, and first-contact provenance was derived from DNS, email, and telephony metadata. For Stages S5/S6 (covert exfiltration), we generated paraphrase-based exfil payloads and protocol-mimic flows seeded from well-known intrusion-detection traces and then embedded them in TLS-like sessions to evaluate network detectors and egress policy.

Appendix E.3. Methodology

Each posture was attacked in twenty-four independent trials. (New cluster, fresh DeepSeek-R1 context, and regenerated synthetic employees per trial.) For every HTTP or gRPC call, the SDP logged the following: timestamp, caller identity, service, decision, and CFKC stage inferred by our labeling model (precision 0.93, recall 0.91 on 1,200 hand-labeled events). Success criteria followed the CFKC phase definitions: progression to a later phase was counted only if all prior phases had been satisfied.

Appendix E.4. Results

Table A2 summarizes the outcomes. Percentages are averaged over the twenty-four trials for each posture.

Table A2. Mean progression and dwell time by CFKC phase (percent of trials, mean ± std).

ZT Posture	Phase Reached (% of Trials)						
	TI	**PP**	**EN**	**DC**	**EX**	**MN**	**CE**
P_{static}	100.0	100.0	91.7	79.2	70.8	45.8	29.2
$P_{adaptive}$	100.0	91.7	41.7	20.8	4.2	0.0	0.0
Dwell time [min]	(mean from initial contact to first definitive detection)						
P_{static}	29.4 ± 6.7						
$P_{adaptive}$	8.9 ± 2.4						

Stage S3 (Engagement): voice resets under first-contact provenance and liveness.

Table A3 reports quantitative effects for the help-desk reset workflow. Combining first-contact provenance with on-device liveness reduced successful voice-reset bypasses from 17.1% to 6.3% with modest latency overhead, while raising operator override precision through short, human-readable explanations.

Table A3. S3 voice-assisted resets: provenance+liveness effects on bypass rate, latency, and operator workflow.

Condition	Bypass Rate (%)	Decision Latency (ms)	Operator Override Precision	N (Reset Attempts)
Baseline (no provenance, no liveness)	17.1 [14.9, 19.4]	84 (p50), 162 (p95)	0.71	1184
Provenance + liveness (on-device)	6.3 [5.0, 7.8]	106 (p50), 189 (p95)	0.83	1176

Stages S5/S6 (Execution/Monetization): tamper-proof monitoring and on-device exfil classifiers.

Table A4 shows network-level metrics for content-aware egress controls. End-to-end verifiable logging paired with on-device exfil classifiers increased detector AUC from 0.811

to 0.914, reduced the median revoke time by roughly half, and shortened the detection horizon from megabytes to near-megabyte scales at comparable true-positive rates.

Table A4. S5/S6 exfiltration controls: NIDS AUC, revoke time, and detection horizon at matched TPR.

Condition	NIDS AUC	FPR at TPR = 0.90	Median Revoke Time (s)	Detection Horizon (MB)
Baseline (central NIDS only)	0.811 [0.802, 0.822]	8.7%	93.8 [86.9, 101.6]	2.63 [2.41, 2.86]
Tamper-proof monitoring + on-device exfil classifiers	0.914 [0.905, 0.922]	5.1%	45.6 [41.8, 50.7]	1.07 [0.96, 1.18]

Appendix E.5. Discussion

Correlation with CFKC predictions.

AutoGFraud invariably completed *Target Identification* and *Planning* (TI, PP) in both postures, consistent with the **high** impact levels assigned to identity governance and context-aware trust algorithms for those stages in Figure 5. Under the static posture, 22/24 trials reached *Engagement*, 19/24 reached *Deception*, and 17/24 reached *Execution*; 11/24 monetised at least $2650 (median $2740) before detection—aligning with the elevated risk predicted for the Execution–Monetization span.

Efficacy of adaptive signals and explainability.

The adaptive posture cut mean dwell time to 8.9 min and blocked every monetization attempt. In S3 (Table A3), first-contact provenance combined with liveness reduced voice-reset bypass from 17.1% to 6.3% with a p50 latency increase of 22 ms and p95 of 27 ms, and improved operator override precision from 0.71 to 0.83, supporting the prioritization of human-actionable explanations and operator-aware workflows in Section 7. The inspection of SDP logs shows that 68.3% of blocked requests were triggered by micro-deviations in typing cadence and time-of-day, reinforcing continuous context signals as a decisive ZT lever against generative-AI fraud.

Monitoring integrity and egress control.

In S5/S6 (Table A4), tamper-proof monitoring and on-device exfil classifiers increased NIDS AUC from 0.811 to 0.914 at a fixed TPR of 0.90, lowered the FPR from 8.7% to 5.1%, and reduced the median revoke time from 93.8 s to 45.6 s, aligning with end-to-end evidence requirements and hardware-rooted enforcement. The detection horizon tightened from 2.63 MB to 1.07 MB, which materially limits the value of partial leaks.

Residual gaps.

Although $P_{adaptive}$ prevented monetary loss, 10/24 trials still reached *Engagement*. Live deep-fake voice calls refined in-session bypassed voice-print MFA in one instance, indicating that behavioral biometrics alone cannot shoulder CFKC phases 3–4 and require redundant provenance controls.

Appendix E.6. Limitations and Future Work

The cluster comprised only 23 microservices and one identity provider; complex B2B federations may yield different trust-boundary leakages. *AutoGFraud* had no reinforcement-learning loop, thus represents a lower bound on what iterative agents could achieve. Future evaluations will incorporate a blue-team co-pilot to study human/AI concurrency—a gap

identified in Section 6.5—and will extend S5/S6 exfil tests to cover non-HTTP transports and mobile-edge gateways.

Appendix E.7. Implications for Zero-Trust Research

The pilot confirms three key insights:

1. **Phase-specific defense is essential.** Blanket controls detected the attack too late; context-adaptive checks blocked progression exactly at CFKC phases where Figure 5 predicts a color shift from *moderate* to *high* impact.
2. **Generative AI magnifies existing ZT blind spots.** All successes exploited Trust-Algorithm weaknesses already catalogued in Section 5, but the adversary reached them faster and stealthier than non-AI baselines in the literature.
3. **Empirical CFKC mapping aids prioritization.** Logging which phase each alert disrupts enables security teams to quantify residual attack surface in CFKC terms—an actionable metric missing from current ZT maturity models.

These observations substantiate the CFKC's explanatory power and provide an empirical anchor for the survey's call—Section 8—to embed generative-AI-aware controls in future ZT architectures.

Appendix F. CFKC Mitigation Playbook and Policy Checklist

This appendix consolidates concrete countermeasures that map the CFKC stages defined in Section 6.2 to the core zero-trust components in NIST SP 800–207, with explicit monitoring signals and privacy-by-design controls. It supplements the workflow in Section 6.3 and operationalizes RQ6 (Section RQ6) by describing where to place evidence collection and how to bound privacy risk while preserving access-control fidelity. We also provide a policy checklist that can be adopted with minimal engineering overhead. Items flagged with * are the high-impact actions referenced in Section 7 (H2, H4, H5). Integration hooks refer to the control flow of Algorithm A1 (Appendix A).

Appendix F.1. CFKC Mitigation Playbook: People, Process, and Technology Alignment

Notation. CFKC stages below are listed as *CFKC-Si* with a short descriptor to assist readers. Use the canonical names from Section 6.2 when integrating into figures and code. NIST SP 800-207 components follow the standard abbreviations: policy enforcement point (PEP), policy decision point (PDP), policy administrator (PA), and policy information point (PIP). "Evidence pipeline" refers to the linked logging, attestation, and anomaly-scoring path discussed in Sections 6.3 and 7.

Operational guidance.

To keep deployment tractable, instantiate the evidence pipeline with append-only logs that are cryptographically chained and time-stamped at the PEP and PDP. Place anomaly scorers on the endpoint or gateway where feasible to reduce centralization of raw telemetry. For audit readiness and to support **H5** in Section 7, persist only signed decision summaries, liveness residuals, and policy digests, with retention windows aligned to legal basis and business need.

Table A5. Mitigation playbook mapping CFKC stages to NIST SP 800-207 components, concrete control changes, monitoring signals, and privacy-by-design measures. Items with * are referenced in Section 7 as high-impact contributions to H2, H4, or H5.

CFKC Stage (as in Section 6.2)	Primary NIST 800-207 Components	Concrete Control Changes (Configuration-Level)	Monitoring Signals and Operational Indicators	Privacy-by-Design Controls	Integration Hook into Algorithm A1
CFKC-S1: Reconnaissance and data staging	PIP, PDP	Enforce *inventory-backed* data-access scopes for discovery services. Require signed catalog queries and cap query burst for unauthenticated discovery endpoints.	Telemetry on catalog query entropy, unauthenticated scan rates, unusual API method mix, and time-window clustering.	On-device feature extraction for scan patterns, retention cap of raw endpoint logs to 7 days, purpose limitation tags on catalog responses.	Use `ContextualInformation` capture before `IdentityVerification`. Evidence forwarded to PDP at line 3–4.
CFKC-S2: Content synthesis and persona creation	PDP, PA	* Require *attested identity binding* for sign-up and role elevation, including possession factors with liveness checks hardened against diffusion-model spoofing. Enforce cooling-off periods for privilege changes.	Biometric liveness residuals, keyboard and pointer micro-drift, text stylometry shift, failed liveness correlation across channels.	Local liveness scoring with discard of raw biometrics post-decision, storage of signed decision summaries only, k-anonymity for stylometry features.	Extend `IdentityVerification` to ingest liveness scores. If below threshold, force `DenyAccess`; otherwise, attach signed proof to `ResourceAccessDecision`.
CFKC-S3: Initial contact and social engineering	PEP, PDP	Conditional access that weights *relationship provenance*: recent first-contact, external domain reputation, and content authenticity hints for voice, video, and text. * Auto-sandbox unknown communication channels.	First-contact flags, deepfake likelihood scores, mismatch between stated and observed channel metadata, operator decision latency.	Client-side pre-filtering of media, strip raw payloads after model inference, consent banners for recording and provenance display.	Insert a `pre-access` check before `IdentityVerification`: high-risk first contacts downgrade policy at PDP and force step-up verification.
CFKC-S4: Credential priming and pressure	PEP, PA	Rate-limit authentication prompts and block *prompt-bombing*. Require proof-of-possession tokens bound to device attestation. * Enforce *phishing-resistant* flows (FIDO2/WebAuthn) for all privileged actions.	Spike in push denials, geovelocity anomalies, device attestation mismatch, token binding failures.	On-device proof-of-possession checks, discard raw device attestations after verification, keep hash-chained summaries only.	Strengthen `IdentityVerification`: if multi-prompt heuristics trip, short-circuit to `DenyAccess` and trigger `NotifySecurityTeam`.
CFKC-S5: Policy evasion and lateral movement	PDP, PEP, PA	* Move policy evaluation into attested TEEs for high-value resources; expose signed policy digests. Dynamic least-privilege recomputation on session drift.	Policy-digest mismatch, unexpected resource graph traversal, anomalous East-West flows, time-to-privilege-escalation.	TEE-resident policy eval with minimal telemetry egress, differential privacy on movement heatmaps, 30-day retention for signed digests.	Bind `ResourceAccessDecision` to TEE-verified policy. If digest verification fails, force `DenyAccess` and log evidence.
CFKC-S6: Data exfiltration and monetization	PEP, PDP	Content-aware egress controls with *purpose tags*. Token-bucket per subject-resource pair. * Real-time *exfil classifiers* on device for known templates and generative paraphrase.	Sudden entropy reduction in exports, paraphrase similarity to sensitive templates, covert channel signatures, breakout to unmanaged sinks.	Redact at source, on-device inference for exfil classifiers, store only policy decisions and hashes of exemplar matches.	Connect `MonitorActivity` to egress detectors. If `ThreatDetected`, call `RevokeAccess` and emit signed egress report.
CFKC-S7: Cleanup and persistence	PA, PIP	Golden-image attestation at session end, privilege decay timers, and drift reconciliation of local policy caches.	Residual scheduled tasks, unexpected service registrations, policy-cache divergence, failed attestation on teardown.	Retain teardown attestations only, purge transient identifiers, rotate keys using short-lived credentials.	After `RevokeAccess` or normal end, require `MonitorActivity` to confirm teardown attestation and close evidence chain.

Appendix F.2. Policy Checklist Derived from the Playbook

Table A6 lists concrete settings that can be rolled out in phases. The checklist emphasizes measurable verification and explicit placement in the zero-trust control plane. Items **#1**, **#4**, **#6**, **#9**, and **#12** are the high-impact actions called out in Section 7 and align with **H2** (tamper-proof, privacy-bounded monitoring), **H4** (policy enclaves with hardware roots of trust), and **H5** (explainable AI-governed GRC). Each row specifies the expected benefit and how to verify it in production or pre-production. Where the checklist references Algorithm A1, the hook indicates the function where the control integrates.

Deployment notes.

Start with Items **1**, **4**, **6**, **9**, and **12** to realize the improvements promised by **H2**, **H4**, and **H5**. Tie each to a verification ritual. For example, require monthly digest attestation drills for Item **6** and quarterly evidence-chain audits for Item **1**. Integrate explanations from Item **9** into the operator console and record operator overrides in the evidence pipeline. Connect teardown attestations from Item **12** to incident reviews to prove session finality.

Link to Algorithm A1.

The checklist assumes the control points exposed in Algorithm A1. Extend `IdentityVerification` with liveness and device binding, wrap `ResourceAccessDecision` with TEE-backed policy evaluation and explanation generation, and enrich `MonitorActivity` with egress and drift detectors whose alerts call `RevokeAccess` and emit signed, privacy-bounded evidence.

Verification and reporting.

For each adopted setting, maintain a small scorecard with three fields: coverage percentage, performance overhead at p95, and detection or prevention lift versus the pre-adoption baseline. Publish the scorecard as part of security governance reporting to satisfy RQ6 concerns concerning auditability and scalability.

Table A6. Zero-trust policy checklist with settings, rationale, expected benefit, verification, and placement. Stars mark the items highlighted in Section 7 (H2/H4/H5).

ID	Policy Setting	Rationale	Expected Benefit	Verification Metric and Method	Placement
1*	Cryptographically chained, write-once decision logs across PEP and PDP with hardware-anchored time stamps	Establish tamper-evident evidence for investigations and automated rescoring, bound by purpose and retention	Faster incident triage and reliable forensics that support continuous assurance under RQ6	Evidence-chain completeness rate, median query latency for incident review, and absence of chain breaks in quarterly audits	PEP, PDP; Algorithm A1 `MonitorActivity`
2	First-contact downgrading with auto-sandbox for unknown external identities and channels	Reduce social engineering success by isolating untrusted flows while collecting provenance	Lower first-contact compromise rate and improved operator confidence	Reduction in first-contact grant decisions without step-up; mean time-to-detection for sandboxed flows	PEP, PDP; pre-`IdentityVerification`
3	Cooling-off periods for role elevation with attested device checks	Prevent rushed privilege changes under attacker pressure	Fewer privilege-escalation incidents and better change auditability	Count of elevation attempts within cooling windows that were blocked; false-positive rate	PA, PDP; `ResourceAccessDecision`
4*	Phishing-resistant authentication for all privileged actions (FIDO2/WebAuthn with device-bound tokens)	Neutralize push fatigue and credential replay amplified by generative content	Step-change reduction in high-impact account takeovers	Rate of push-denial spikes; takeover incident rate; coverage percentage for FIDO across privileged roles	PEP; `IdentityVerification`
5	Relationship-provenance weighting in access scoring (recency, domain reputation, channel authenticity)	Make access risk-aware to social and organizational context	Lower acceptance of spoofed relationships and channels	AUC lift of access-scoring model with provenance features vs. baseline; approval reversal rate after manual review	PDP; `ResourceAccessDecision`
6*	Policy evaluation in attested TEEs with published signed policy digests	Remove policy-tampering avenues and enable verifiable conformance	Stronger assurance for high-value resources and faster attestable audits	Share of decisions produced in TEEs; digest verification failure rate; p95 latency overhead	PDP; `ResourceAccessDecision`
7	Dynamic least-privilege recomputation on session drift (risk score, resource graph changes)	Constrain lateral movement accelerated by automated tooling	Lower dwell time and smaller blast radius	Median time-to-privilege-reduction after drift; percentage of sessions auto-downgraded	PDP, PEP; `MonitorActivity`
8	Egress token-bucket per subject-resource pair with content-aware classifiers on device	Throttle and detect exfiltration including paraphrase and format-shift	Fewer large-scale exfiltration events and better near-real-time containment	Number of blocked egress bursts; detection rate on holdout exfil templates; median revoke time	PEP; `MonitorActivity` and `RevokeAccess`
9*	Evidence schemas and audit grammar that map decisions to NIST SP 800–207 objectives with human-actionable explanations	Improve explainability for denials and grants and standardize audits across teams	Faster exception handling and fewer misconfigurations	Decision-explanation coverage, operator resolution time, audit exception rate	PDP, PA; `ResourceAccessDecision`
10	On-device liveness scoring and immediate discard of raw biometrics after decision	Reduce privacy risk while increasing resistance to synthetic identity	Improved privacy posture without loss in liveness accuracy	False-accept and false-reject rates; retention audit of biometric artifacts	PEP; `IdentityVerification`
11	Query-burst caps and signed catalog requests for discovery endpoints	Limit unauthenticated reconnaissance and data staging	Reduction in scanning footprint and better early-stage detection	Unauthenticated scan rate, catalog query entropy distributions	PIP, PDP; `ContextualInformation`
12*	Attested teardown with privilege decay timers and cache-drift reconciliation	Ensure sessions end cleanly and remove persistence footholds	Lower rate of residual privileges and shadow tasks	Percentage of sessions with verified teardown; divergence between local and central policy caches	PA, PEP; `MonitorActivity`
13	Privacy budgets and retention caps for telemetry with summarization at source	Bound data collection while maintaining detection efficacy	Reduced data exposure without material drop in detection	Detection AUC vs. privacy budget; retained-bytes trend vs. baseline	Evidence pipeline; cross-cuts PEP/PDP
14	Redaction of sensitive payloads prior to central processing with hash-chained summaries	Minimize central storage of raw content while preserving verifiability	Lower breach impact and simpler lawful basis management	Share of events with redacted payloads; verification success using summaries	Evidence pipeline; PEP
15*	Continuous conformance tests using synthetic attack traces covering all CFKC stages	Keep policies fresh against new model exploits and regressions	Earlier detection of policy drift and exploitable gaps	Time-to-fail discovery after policy change; number of regressions caught pre-production	PA, PDP; CI/CD for policy; informs `ResourceAccessDecision`

References

1. Syed, N.F.; Shah, S.W.; Shaghaghi, A.; Anwar, A.; Baig, Z.; Doss, R. ZTA: A comprehensive survey. *IEEE Access* **2022**, *10*, 57143–57179. [CrossRef]
2. Azad, M.A.; Abdullah, S.; Arshad, J.; Lallie, H.; Ahmed, Y.H. Verify and trust: A multidimensional survey of zero-trust security in the age of IoT. *Internet Things* **2024**, *27*, 101227. [CrossRef]
3. Gupta, A.; Gupta, R.; Jadav, D.; Tanwar, S.; Kumar, N.; Shabaz, M. Proxy smart contracts for zero trust architecture implementation in Decentralised Oracle Networks based applications. *Comput. Commun.* **2023**, *206*, 10–21. [CrossRef]
4. Rose, S.; Borchert, O.; Mitchell, S.; Connelly, S. Zero Trust Architecture. In *NIST Special Publication*; National Institute of Standards and Technology: Gaithersburg, MD, USA, 2020. [CrossRef]
5. Chen, H.; Babar, M.A. Security for Machine Learning-based Software Systems: A Survey of Threats, Practices, and Challenges. *ACM Comput. Surv.* **2024**, *56*, 1–38. [CrossRef]
6. McIntosh, T.; Susnjak, T.; Liu, T.; Xu, D.; Watters, P.; Liu, D.; Hao, Y.; Ng, A.; Halgamuge, M. Ransomware reloaded: Re-examining its trend, research and mitigation in the era of data exfiltration. *ACM Comput. Surv.* **2024**, *57*, 1–40 . [CrossRef]
7. Nahar, N.; Andersson, K.; Schelén, O.; Saguna, S. A Survey on Zero Trust Architecture: Applications and Challenges of 6G Networks. *IEEE Access* **2024**, *12*, 94753–94764. [CrossRef]
8. Itodo, C.; Ozer, M. Multivocal Literature Review on Zero-Trust Security Implementation. *Comput. Secur.* **2024**, *141*, 103827. [CrossRef]
9. Buck, C.; Olenberger, C.; Schweizer, A.; Völter, F.; Eymann, T. Never trust, always verify: A multivocal literature review on current knowledge and research gaps of zero-trust. *Comput. Secur.* **2021**, *110*, 102436. [CrossRef]
10. Sarkar, S.; Choudhary, G.; Shandilya, S.K.; Hussain, A.; Kim, H. Security of zero trust networks in cloud computing: A comparative review. *Sustainability* **2022**, *14*, 11213. [CrossRef]
11. Yan, X.; Wang, H. Survey on zero-trust network security. In Proceedings of the Artificial Intelligence and Security: 6th International Conference, ICAIS 2020, Hohhot, China, 17–20 July 2020; Proceedings, Part I 6; Springer: Berlin/Heidelberg, Germany, 2020; pp. 50–60.
12. Kang, H.; Liu, G.; Wang, Q.; Meng, L.; Liu, J. Theory and application of zero trust security: A brief survey. *Entropy* **2023**, *25*, 1595. [CrossRef]
13. Dhiman, P.; Saini, N.; Gulzar, Y.; Turaev, S.; Kaur, A.; Nisa, K.U.; Hamid, Y. A Review and Comparative Analysis of Relevant Approaches of Zero Trust Network Model. *Sensors* **2024**, *24*, 1328. [CrossRef] [PubMed]
14. He, Y.; Huang, D.; Chen, L.; Ni, Y.; Ma, X. A survey on zero trust architecture: Challenges and future trends. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 6476274. [CrossRef]
15. Stafford, V. Zero trust architecture. In *NIST Special Publication*; National Institute of Standards and Technology: Gaithersburg, MD, USA, 2020.
16. Kindervag, J. Build security into your network's dna: The zero trust network architecture. *Forrester Res. Inc.* **2010**, *27*, 1–16.
17. Ward, R.; Beyer, B. Beyondcorp: A new approach to enterprise security. *Mag. Usenix Sage* **2014**, *39*, 6–11.
18. McIntosh, T.; Liu, T.; Susnjak, T.; Alavizadeh, H.; Ng, A.; Nowrozy, R.; Watters, P. Harnessing GPT-4 for generation of cybersecurity GRC policies: A focus on ransomware attack mitigation. *Comput. Secur.* **2023**, *134*, 103424. [CrossRef]
19. Janssen, M.; Brous, P.; Estevez, E.; Barbosa, L.S.; Janowski, T. Data governance: Organizing data for trustworthy Artificial Intelligence. *Gov. Inf. Q.* **2020**, *37*, 101493. [CrossRef]
20. Kampa, T.; Müller, C.K.; Großmann, D. Interlocking IT/OT security for edge cloud-enabled manufacturing. *Ad Hoc Netw.* **2024**, *154*, 103384. [CrossRef]
21. Paya, A.; Gómez, A. Securesdp: A novel software-defined perimeter implementation for enhanced network security and scalability. *Int. J. Inf. Secur.* **2024**, *23*, 2793–2808. [CrossRef]
22. Alevizos, L.; Ta, V.T.; Hashem Eiza, M. Augmenting zero trust architecture to endpoints using blockchain: A state-of-the-art review. *Secur. Priv.* **2022**, *5*, e191. [CrossRef]
23. Bahmani, R.; Brasser, F.; Dessouky, G.; Jauernig, P.; Klimmek, M.; Sadeghi, A.R.; Stapf, E. CURE: A Security Architecture with Customizable and Resilient Enclaves. In Proceedings of the 30th USENIX Security Symposium (USENIX Security 2021), Virtual Conference, 11–13 August 2021; pp. 1073–1090.
24. Xu, K.; Chen, M.; Yue, S.; Zhang, F.; Wang, J.; Wen, Y.; Lü, G. The portal of OpenGMS: Bridging the contributors and users of geographic simulation resources. *Environ. Model. Softw.* **2024**, *180*, 106142. [CrossRef]
25. McIntosh, T.; Kayes, A.; Chen, Y.P.; Ng, A.; Watters, P. Ransomware mitigation in the modern era: A comprehensive review, research challenges, and future directions. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–16. [CrossRef]
26. Ali, B.; Hijjawi, S.; Campbell, L.H.; Gregory, M.A.; Li, S. A maturity framework for zero-trust security in multiaccess edge computing. *Secur. Commun. Netw.* **2022**, *2022*, 3178760. [CrossRef]
27. Dekker, M.; Alevizos, L. A threat-intelligence driven methodology to incorporate uncertainty in cyber risk analysis and enhance decision-making. *Secur. Priv.* **2024**, *7*, e333. [CrossRef]

28. Al Shehhi, F.; Otoum, S. On the Feasibility of Zero-Trust Architecture in Assuring Security in Metaverse. In Proceedings of the 2023 International Conference on Intelligent Metaverse Technologies & Applications (iMETA), Tartu, Estonia, 18–20 September 2023; pp. 1–8.
29. McIntosh, T.; Kayes, A.; Chen, Y.P.P.; Ng, A.; Watters, P. Dynamic user-centric access control for detection of ransomware attacks. *Comput. Secur.* **2021**, *111*, 102461. [CrossRef]
30. McIntosh, T.; Kayes, A.; Chen, Y.P.; Ng, A.; Watters, P. Applying staged event-driven access control to combat ransomware. *Comput. Secur.* **2023**, *128*, 103160. [CrossRef]
31. Ometov, A.; Petrov, V.; Bezzateev, S.; Andreev, S.; Koucheryavy, Y.; Gerla, M. Challenges of multi-factor authentication for securing advanced IoT applications. *IEEE Netw.* **2019**, *33*, 82–88. [CrossRef]
32. Habbal, A.; Ali, M.K.; Abuzaraida, M.A. Artificial Intelligence Trust, risk and security management (AI trism): Frameworks, applications, challenges and future research directions. *Expert Syst. Appl.* **2024**, *240*, 122442. [CrossRef]
33. Simpson, W.R.; Foltz, K.E. Network Segmentation and Zero Trust Architectures. In Proceedings of the Lecture Notes in Engineering and Computer Science, Proceedings of the World Congress on Engineering (WCE 2021), London, UK, 7–9 July 2021; pp. 201–206.
34. Bello, Y.; Hussein, A.R.; Ulema, M.; Koilpillai, J. On sustained zero trust conceptualization security for mobile core networks in 5g and beyond. *IEEE Trans. Netw. Serv. Manag.* **2022**, *19*, 1876–1889. [CrossRef]
35. Ramezanpour, K.; Jagannath, J. Intelligent zero trust architecture for 5G/6G networks: Principles, challenges, and the role of machine learning in the context of O-RAN. *Comput. Netw.* **2022**, *217*, 109358. [CrossRef]
36. Gudala, L.; Shaik, M.; Venkataramanan, S. Leveraging Machine Learning for Enhanced Threat Detection and Response in Zero Trust Security Frameworks: An Exploration of Real-Time Anomaly Identification and Adaptive Mitigation Strategies. *J. Artif. Intell. Res.* **2021**, *1*, 19–45.
37. Khamvilai, T.; Pakmehr, M. Zero Trust Avionics Systems (ZTAS). In Proceedings of the 2023 IEEE/AIAA 42nd Digital Avionics Systems Conference (DASC 2023), Barcelona, Spain, 1–5 October 2023; pp. 1–8.
38. Zanasi, C.; Russo, S.; Colajanni, M. Flexible zero trust architecture for the cybersecurity of industrial IoT infrastructures. *Ad Hoc Netw.* **2024**, *156*, 103414. [CrossRef]
39. Fernandez, E.B.; Brazhuk, A. A critical analysis of ZTA. *Comput. Stand. Interfaces* **2024**, *89*, 103832. [CrossRef]
40. Edo, O.C.; Ang, D.; Billakota, P.; Ho, J.C. A zero trust architecture for health information systems. *Health Technol.* **2024**, *14*, 189–199. [CrossRef]
41. Rivera, J.J.D.; Muhammad, A.; Song, W.C. Securing Digital Identity in the Zero Trust Architecture: A Blockchain Approach to Privacy-Focused Multi-Factor Authentication. *IEEE Open J. Commun. Soc.* **2024**, *5*, 2792–2814. [CrossRef]
42. Chaturvedi, I.; Pawar, P.M.; Muthalagu, R.; Tamizharasan, P. Zero Trust Security Architecture for Digital Privacy in Healthcare. In *Information Technology Security: Modern Trends and Challenges*; Springer Nature: Singapore, 2024; pp. 1–23.
43. Colomb, Y.; White, P.; Islam, R.; Alsadoon, A. Applying Zero Trust Architecture and Probability-Based Authentication to Preserve Security and Privacy of Data in the Cloud. In *Emerging Trends in Cybersecurity Applications*; Springer International Publishing: Cham, Switzerland, 2022; pp. 137–169.
44. ElSayed, Z.; Elsayed, N.; Bay, S. A Novel Zero-Trust Machine Learning Green Architecture for Healthcare IoT Cybersecurity: Review, Analysis, and Implementation. *SoutheastCon* **2024**, *2024*, 686–692.
45. Barzyk, C.; Hickson, J.; Ochoa, J.; Talley, J.; Willeke, M.; Coffey, S.; Pavlik, J.; Bastian, N.D. A Generative Artificial Intelligence Methodology for Automated Zero-Shot Data Tagging to Support Tactical Zero Trust Architecture Implementation. *Ind. Syst. Eng. Rev.* **2025**, 12, *83–88. [CrossRef]*
46. Patil, K.; Desai, B.; Mehta, I.; Patil, A. A Contemporary Approach: Zero Trust Architecture for Cloud-Based Fintech Services. *Innov. Comput. Sci. J.* **2023**, *9*.
47. Stanojevic, M.; Capko, D.; Lendak, I.; Stoja, S.; Jelacic, B. Fighting Insider Threats, with Zero-Trust in Microservice-based, Smart Grid OT Systems. *Acta Polytech. Hung.* **2023**, *20*, 229–248. [CrossRef]
48. Leahy, D.; Thorpe, C. Zero trust container architecture (ztca): A framework for applying zero trust principals to docker containers. In Proceedings of the International Conference on Cyber Warfare and Security, Islamabad, Pakistan, 7–8 December 2022; Volume 17, pp. 111–120.
49. Tsai, M.; Lee, S.; Shieh, S.W. Strategy for implementing of zero trust architecture. *IEEE Trans. Reliab.* **2024**, *73*, 93–100. [CrossRef]
50. Din, I.U.; Khan, K.H.; Almogren, A.; Zareei, M.; Díaz, J.A.P. Securing the Metaverse: A Blockchain-Enabled Zero-Trust Architecture for Virtual Environments. *IEEE Access* **2024**, *12*, 92337–92347. [CrossRef]
51. Ahmadi, S. Zero trust architecture in cloud networks: Application, challenges and future opportunities. *J. Eng. Res. Rep.* **2024**, *26*, 215–228. [CrossRef]
52. Hussain, M.; Pal, S.; Jadidi, Z.; Foo, E.; Kanhere, S. Federated Zero Trust Architecture using Artificial Intelligence. *IEEE Wirel. Commun.* **2024**, *31*, 30–35. [CrossRef]

53. Liu, Y.; Su, Z.; Peng, H.; Xiang, Y.; Wang, W.; Li, R. Zero Trust-Based Mobile Network Security Architecture. *IEEE Wirel. Commun.* **2024**, *31*, 82–88. [CrossRef]
54. Chang, Y.C.; Lin, Y.S.; Sangaiahc, A.K.; Wu, H.T. A Private Blockchain System based on Zero Trust Architecture. In Proceedings of the 2024 26th International Conference on Advanced Communications Technology (ICACT), Pyeong Chang, Republic of Korea, 4–7 February 2024; pp. 143–146.
55. Huber, B.; Kandah, F. Zero Trust+: A Trusted-based Zero Trust architecture for IoT at Scale. In Proceedings of the 2024 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 6–8 January 2024; pp. 1–6.
56. Zhang, H.; Wang, Q.; Zhang, X.; He, Y.; Tang, B.; Li, Q. Toward Zero-Trust IoT Networks via Per-Packet Authorization. *IEEE Commun. Mag.* **2024**, *62*, 90–96. [CrossRef]
57. Dhanapal, A.D.; Ramanujan, S.A.; Jeyalakshmi, V. Trust-Free Homes: The Zero-Trust Paradigm in a Smart Home Setting. In *Communication Technologies and Security Challenges in IoT: Present and Future*; Springer: Singapore, 2024; pp. 335–349.
58. Guleri, A.; Singh, N.P.; Singh, P.; Lata, K. Siddu: Decentralized Authorization with Zero Trust. In Proceedings of the International Conference on Communications and Cyber Physical Engineering 2018, Hyderabad, India, 28–29 February 2024; pp. 213–221.
59. Yang, Y.; Bai, F.; Yu, Z.; Shen, T.; Liu, Y.; Gong, B. An anonymous and supervisory cross-chain privacy protection protocol for zero-trust IoT application. *ACM Trans. Sens. Netw.* **2024**, *20*, 1–20. [CrossRef]
60. Jamil, M.; Farhan, M.; Ullah, F.; Srivastava, G. A Lightweight Zero Trust Framework for Secure 5G VANET Vehicular Communication. *IEEE Wirel. Commun.* **2024**, *31*, 136–141. [CrossRef]
61. Xu, X.; Zhou, X.; Zhou, X.; Bilal, M.; Qi, L.; Xia, X.; Dou, W. Distributed Edge Caching for Zero Trust-Enabled Connected and Automated Vehicles: A Multi-Agent Reinforcement Learning Approach. *IEEE Wirel. Commun.* **2024**, *31*, 36–41. [CrossRef]
62. Nawshin, F.; Unal, D.; Hammoudeh, M.; Suganthan, P.N. AI-powered malware detection with Differential Privacy for zero trust security in Internet of Things networks. *Ad Hoc Netw.* **2024**, *161*, 103523. [CrossRef]
63. Nkoro, E.C.; Njoku, J.N.; Nwakanma, C.I.; Lee, J.M.; Kim, D.S. Zero-Trust Marine Cyberdefense for IoT-Based Communications: An Explainable Approach. *Electronics* **2024**, *13*, 276. [CrossRef]
64. Farouk, A.; Al-Kuwari, S.; Abulkasim, H.; Mumtaz, S.; Adil, M.; Song, H. Quantum Computing: A Tool for Zero-trust Wireless Networks. *IEEE Netw.* **2024**, *39*, 140–148. [CrossRef]
65. Javeed, D.; Saeed, M.S.; Adil, M.; Kumar, P.; Jolfaei, A. A federated learning-based zero trust intrusion detection system for Internet of Things. *Ad. Hoc. Netw.* **2024**, *162*, 103540. [CrossRef]
66. Liu, C.; Tan, R.; Wu, Y.; Feng, Y.; Jin, Z.; Zhang, F.; Liu, Y.; Liu, Q. Dissecting zero trust: Research landscape and its implementation in IoT. *Cybersecurity* **2024**, *7*, 20. [CrossRef]
67. Fang, H.; Zhu, Y.; Zhang, Y.; Wang, X. Decentralized Edge Collaboration for Seamless Handover Authentication in Zero-Trust IoV. *IEEE Trans. Wirel. Commun.* **2024**, *23*, 8760–8772. [CrossRef]
68. Okegbile, S.D.; Cai, J.; Chen, J.; Yi, C. A Reputation-Enhanced Shard-Based Byzantine Fault-Tolerant Scheme for Secure Data Sharing in Zero Trust Human Digital Twin Systems. *IEEE Internet Things J.* **2024**, *11*, 22726–22741. [CrossRef]
69. Al Shahrani, A.M.; Rizwan, A.; Sánchez-Chero, M.; Cornejo, L.L.C.; Shabaz, M. Blockchain-enabled federated learning for prevention of power terminals threats in IoT environment using edge zero-trust model. *J. Supercomput.* **2024**, *80*, 7849–7875. [CrossRef]
70. Rasool, S.; Saleem, A.; ul Haq, M.I.; Jacobsen, R.H. Towards Zero Trust Security for Prosumer-Driven Verifiable Green Energy Certificates. In Proceedings of the 2024 7th International Conference on Energy Conservation and Efficiency (ICECE), Lahore, Pakistan, 6–7 March 2024; pp. 1–6.
71. Cao, H.; Yang, L.; Garg, S.; Alrashoud, M.; Guizani, M. Softwarized resource allocation of tailored services with zero security trust in 6G networks. *IEEE Wirel. Commun.* **2024**, *31*, 58–65. [CrossRef]
72. Sullivan, B.; Khan, J.A. OBSERVE: Blockchain-Based Zero Trust Security Protocol for Connected and Autonomous Vehicles (CAVs) Data Using Simple Machine Learning. In Proceedings of the 2024 International Conference on Computing, Networking and Communications (ICNC), Big Island, HI, USA, 19–22 February 2024; pp. 554–559.
73. Trott, D. A zero-trust journey through the threat landscape. *Netw. Secur.* **2024**, *2024*, 2. [CrossRef]
74. Zhu, L.; Huang, D.; Na, Y.; Li, X. Design and Stability Analysis of Vehicle Platooning Control in Zero-Trust Environment. In Proceedings of the 2024 IEEE 13th Data Driven Control and Learning Systems Conference (DDCLS), Kaifeng, China, 17–19 May 2024; pp. 2076–2081.
75. Nakamura, S.; Takizawa, M. Trust zone model with the mandatory access control model. In Proceedings of the International Conference on Emerging Internet, Data & Web Technologies, Naples, Italy, 21–23 February 2024; pp. 512–521.
76. Awan, K.A.; Din, I.U.; Almogren, A.; Kim, B.S.; Guizani, M. Enhancing IoT Security with Trust Management Using Ensemble XGBoost and AdaBoost Techniques. *IEEE Access* **2024**, *12*, 116609–116621. [CrossRef]
77. Kumar, R.; Aljuhani, A.; Javeed, D.; Kumar, P.; Islam, S.; Islam, A.N. Digital twins-enabled zero touch network: A smart contract and explainable AI integrated cybersecurity framework. *Future Gener. Comput. Syst.* **2024**, *156*, 191–205. [CrossRef]

78. Khan, M.J. Zero trust architecture: Redefining network security paradigms in the digital age. *World J. Adv. Res. Rev.* **2023**, *19*, 105–116. [CrossRef]
79. Federici, F.; Martintoni, D.; Senni, V. A zero-trust architecture for remote access in industrial IoT infrastructures. *Electronics* **2023**, *12*, 566.
80. Chen, X.; Feng, W.; Ge, N.; Zhang, Y. Zero trust architecture for 6G security. *IEEE Netw.* **2023**, *38*, 224–232. [CrossRef]
81. Anderson, J.; Huang, Q.; Cheng, L.; Hu, H. A Zero Trust Architecture for Connected and Autonomous Vehicles. *IEEE Internet Comput.* **2023**, *27*, 7–14. [CrossRef]
82. Bradatsch, L.; Miroshkin, O.; Kargl, F. ZTSFC: A Service Function Chaining-Enabled Zero Trust Architecture. *IEEE Access* **2023**, *11*, 125307–125327. [CrossRef]
83. Wang, Z.; Jin, M.; Jiang, L.; Feng, C.; Cao, J.; Yun, Z. Secure access method of power internet of things based on zero trust architecture. In Proceedings of the International Conference on Swarm Intelligence, Shenzhen, China, 14–18 July 2023; pp. 386–399.
84. Sedjelmaci, H.; Ansari, N. Zero trust architecture empowered attack detection framework to secure 6g edge computing. *IEEE Netw.* **2023**, *38*, 196–202. [CrossRef]
85. Feng, X.; Hu, S. Cyber-physical zero trust architecture for industrial cyber-physical systems. *IEEE Trans. Ind.-Cyber-Phys. Syst.* **2023**, *1*, 394–405. [CrossRef]
86. Xu, M.; Guo, J.; Yuan, H.; Yang, X. Zero-Trust Security Authentication Based on SPA and Endogenous Security Architecture. *Electronics* **2023**, *12*, 782. [CrossRef]
87. Hong, S.; Xu, L.; Huang, J.; Li, H.; Hu, H.; Gu, G. SysFlow: Toward a programmable zero trust framework for system security. *IEEE Trans. Inf. Forensics Secur.* **2023**, *18*, 2794–2809. [CrossRef]
88. Kholidy, H.A.; Disen, K.; Karam, A.; Benkhelifa, E.; Rahman, M.A.; Rahman, A.U.; Almazyad, I.; Sayed, A.F.; Jaziri, R. Secure the 5G and beyond networks with zero trust and access control systems for cloud native architectures. In Proceedings of the 2023 20th ACS/IEEE International Conference on Computer Systems and Applications (AICCSA), Giza, Egypt, 4–7 December 2023; pp. 1–8.
89. Wang, J.; Chen, J.; Xiong, N.; Alfarraj, O.; Tolba, A.; Ren, Y. S-BDS: An effective blockchain-based data storage scheme in zero-trust IoT. *ACM Trans. Internet Technol.* **2023**, *23*, 1–23. [CrossRef]
90. Tanque, M.; Foxwell, H.J. Cyber risks on IoT platforms and zero trust solutions. In *Advances in Computers*; Elsevier: Amsterdam, The Netherlands, 2023; Volume 131, pp. 79–148.
91. Awale, V.; Gaikwad, S. Zero Trust Architecture Using Hyperledger Fabric. In Proceedings of the 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 6–8 July 2023; pp. 1–4.
92. Huang, W.; Xie, X.; Wang, Z.; Feng, J.; Han, G.; Zhang, W. ZT-Access: A combining zero trust access control with attribute-based encryption scheme against compromised devices in power IoT environments. *Ad Hoc Netw.* **2023**, *145*, 103161. [CrossRef]
93. Che, K.; Sheng, S. Cloud Native Network Security Architecture Strategy under Zero Trust Scenario. In Proceedings of the 2023 IEEE 7th Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing, China, 15–17 September 2023; Volume 7, pp. 867–871.
94. Feng, Y.; Zhong, Z.; Sun, X.; Wang, L.; Lu, Y.; Zhu, Y. Blockchain enabled zero trust based authentication scheme for railway communication networks. *J. Cloud Comput.* **2023**, *12*, 62. [CrossRef]
95. Dong, C.; Pal, S.; An, Q.; Yao, A.; Jiang, F.; Xu, Z.; Li, J.; Lu, M.; Song, Y.; Chen, S.; et al. Securing Smart UAV Delivery Systems Using Zero Trust Principle-Driven Blockchain Architecture. In Proceedings of the 2023 IEEE International Conference on Blockchain (Blockchain), Hainan, China, 17–21 December 2023; pp. 315–322.
96. Saleem, M.; Warsi, M.; Islam, S. Secure information processing for multimedia forensics using zero-trust security model for large scale data analytics in SaaS cloud computing environment. *J. Inf. Secur. Appl.* **2023**, *72*, 103389. [CrossRef]
97. Cheng, R.; Chen, S.; Han, B. Toward zero-trust security for the metaverse. *IEEE Commun. Mag.* **2023**, *62*, 156–162. [CrossRef]
98. Wu, K.; Cheng, R.; Xu, H.; Tong, J. Design and Implementation of the Zero Trust Model in the Power Internet of Things. *Int. Trans. Electr. Energy Syst.* **2023**, *2023*, 6545323. [CrossRef]
99. Wang, Z.; Yu, X.; Xue, P.; Qu, Y.; Ju, L. Research on medical security system based on zero trust. *Sensors* **2023**, *23*, 3774. [CrossRef]
100. N'goran, K.R.; Brou, A.P.B.; Pandry, K.G.; Tetchueng, J.L.; Kermarrec, Y.; Asseu, O. Zero Trust Security Strategy for Collaboration Systems. In Proceedings of the 2023 International Symposium on Networks, Computers and Communications (ISNCC), Doha, Qatar, 23–26 October 2023; pp. 1–6.
101. TN, N.; Pramod, D.; Singh, R. Zero trust security model: Defining new boundaries to organizational network. In Proceedings of the 2023 Fifteenth International Conference on Contemporary Computing, Noida, India, 3–5 August 2023; pp. 603–609.
102. Wang, J.; Wang, Z.; Song, J.; Cheng, H. Attribute and User Trust Score-Based Zero Trust Access Control Model in IoV. *Electronics* **2023**, *12*, 4825. [CrossRef]
103. Ishihara, A.K.; Abdelbaky, M.; Shetye, S. Zero-Trust Architecture for Autonomous Edge Computing. In Proceedings of the Scitech 2023, Moscow, Russia, 28 November–1 December 2023.

104. Mohseni Ejiyeh, A. Real-Time Lightweight Cloud-Based Access Control for Wearable IoT Devices: A Zero Trust Protocol. In Proceedings of the First International Workshop on Security and Privacy of Sensing Systems, Istanbul, Turkiye, 12–17 November 2023; pp. 22–29.
105. Kobayashi, N. Zero Trust Security Framework for IoT Actuators. In Proceedings of the 2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC), Torino, Italy, 27–29 June 2023; pp. 1285–1292.
106. Ge, Y.; Zhu, Q. Gazeta: Game-theoretic zero-trust authentication for defense against lateral movement in 5g iot networks. *IEEE Trans. Inf. Forensics Secur.* **2023**, *19*, 540–554. [CrossRef]
107. Jiang, H.; Chang, H.; Mukherjee, S.; Van der Merwe, J. OZTrust: An O-RAN Zero-Trust Security System. In Proceedings of the 2023 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN), Dresden, Germany, 7–9 November 2023; pp. 129–134.
108. Nwoyibe, O.I.; Philip, O.C.; Odi, A.C. Deployment of Zero Trust Access Security Model for Protection Against Cyber Crimes. *IUP J. Comput. Sci.* **2023**, *17*, 52–59.
109. Anderson, J.; Huang, Q.; Cheng, L.; Hu, H. BYOZ: Protecting BYOD through zero trust network security. In Proceedings of the 2022 IEEE International Conference on Networking, Architecture and Storage (NAS), Philadelphia, PA, USA, 3–4 October 2022; pp. 1–8.
110. Wang, L.; Ma, H.; Li, Z.; Pei, J.; Hu, T.; Zhang, J. A data plane security model of SR-BE/TE based on zero-trust architecture. *Sci. Rep.* **2022**, *12*, 20612. [CrossRef]
111. Ameer, S.; Gupta, M.; Bhatt, S.; Sandhu, R. Bluesky: Towards convergence of zero trust principles and score-based authorization for iot enabled smart systems. In Proceedings of the 27th ACM on Symposium on Access Control Models and Technologies, New York, NY, USA, 8–10 June 2022; pp. 235–244.
112. Li, D.; Zhang, E.; Lei, M.; Song, C. Zero trust in edge computing environment: A blockchain based practical scheme. *Math. Biosci. Eng.* **2022**, *19*, 4196–4216. [CrossRef]
113. Liu, Y.; Hao, X.; Ren, W.; Xiong, R.; Zhu, T.; Choo, K.K.R.; Min, G. A blockchain-based decentralized, fair and authenticated information sharing scheme in zero trust internet-of-things. *IEEE Trans. Comput.* **2022**, *72*, 501–512. [CrossRef]
114. Bandara, E.; Liang, X.; Shetty, S.; Mukkamala, R.; Rahman, A.; Keong, N.W. Skunk-A blockchain and zero trust security enabled federated learning platform for 5G/6G network slicing. In Proceedings of the 2022 19th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), Virtual, 20–23 September 2022; pp. 109–117.
115. Dhanaraj, R.K.; Singh, A.; Nayyar, A. Matyas–Meyer Oseas based device profiling for anomaly detection via deep reinforcement learning (MMODPAD-DRL) in zero trust security network. *Computing* **2024**, *106*, 1933–1962. [CrossRef]
116. Ouiazzane, S.; Addou, M.; Barramou, F. A Zero-Trust Model for Intrusion Detection in Drone Networks. *Int. J. Adv. Comput. Sci. Appl* **2023**, *14*, 525–537. [CrossRef]
117. Pontes, D.; Silva, F.; Falcão, E.; Brito, A. Attesting AMD SEV-SNP Virtual Machines with SPIRE. In Proceedings of the 12th Latin-American Symposium on Dependable and Secure Computing, La Paz, Bolivia, 16–18 October 2023; pp. 1–10.
118. Repetto, M.; Carrega, A.; Rapuzzi, R. An architecture to manage security operations for digital service chains. *Future Gener. Comput. Syst.* **2021**, *115*, 251–266. [CrossRef]
119. Zhang, J.; Zheng, J.; Zhang, Z.; Chen, T.; Qiu, K.; Zhang, Q.; Li, Y. Hybrid isolation model for device application sandboxing deployment in Zero Trust architecture. *Int. J. Intell. Syst.* **2022**, *37*, 11167–11187. [CrossRef]
120. Kroculick, J.B. Zero trust decision analysis for next generation networks. In Proceedings of the Disruptive Technologies in Information Sciences VIII, National Harbor, MA, USA, 21–25 April 2024; Volume 13058, pp. 278–286.
121. Xie, H.; Wang, Y.; Ding, Y.; Yang, C.; Liang, H.; Qin, B. Industrial Wireless Internet Zero Trust Model: Zero Trust Meets Dynamic Federated Learning with Blockchain. *IEEE Wirel. Commun.* **2024**, *31*, 22–29. [CrossRef]
122. Zhang, Q.Y.; Wu, G.R.; Yang, R.; Chen, J.Y. Digital image copyright protection method based on blockchain and zero trust mechanism. *Multimed. Tools Appl.* **2024**, *83*, 77267–77302. [CrossRef]
123. Heino, J.; Jalio, C.; Hakkala, A.; Virtanen, S. JAPPI: An unsupervised endpoint application identification methodology for improved Zero Trust models, risk score calculations and threat detection. *Comput. Netw.* **2024**, *250*, 110606. [CrossRef]
124. Raheman, F. Formulating and Supporting a Hypothesis to Address a Catch-22 Situation in 6G Communication Networks. *J. Inf. Secur.* **2024**, *15*, 340–354. [CrossRef]
125. John, J.; John Singh, K. Trust value evaluation of cloud service providers using fuzzy inference based analytical process. *Sci. Rep.* **2024**, *14*, 18028. [CrossRef]
126. Yeoh, W.; Liu, M.; Shore, M.; Jiang, F. Zero trust cybersecurity: Critical success factors and A maturity assessment framework. *Comput. Secur.* **2023**, *133*, 103412. [CrossRef]
127. Ali, B.; Gregory, M.A.; Li, S. Trust-aware task load balancing in multi-access edge computing based on blockchain and a zero trust security capability framework. *Trans. Emerg. Telecommun. Technol.* **2023**, *34*, e4845. [CrossRef]
128. Alalmaie, A.Z.; Nanda, P.; He, X. ZT-NIDS: Zero Trust, Network Intrusion Detection System. In Proceedings of the SECRYPT, Rome, Italy, 10-12 July 2023; pp. 99–110.

129. Lv, Z.; Chen, C.; Zhang, Z.; Di, L.; Li, N. Zero-Trust Security Protection Architecture for Power Grid Based on FAHP Algorithm. In Proceedings of the 2nd International Conference on Internet of Things, Communication and Intelligent Technology, Xuzhou, China, 22–24 September 2023; pp. 49–61.
130. Park, U.H.; Hong, J.h.; Kim, A.; Son, K.H. Endpoint Device Risk-Scoring Algorithm Proposal for Zero Trust. *Electronics* **2023**, *12*, 1906. [CrossRef]
131. Yunanto, W.; Pao, H.K. User behavior Risk Evaluation in Zero Trust Architecture Environment. In Proceedings of the 2022 IEEE 8th World Forum on Internet of Things (WF-IoT), Yokohama, Japan, 26 October–11 November 2022; pp. 1–6.
132. Pittman, J.M.; Alaee, S.; Crosby, C.; Honey, T.; Schaefer, G.M. Towards a model for zero trust data. *Am. J. Sci. Eng.* **2022**, *3*, 18–24. [CrossRef]
133. García-Teodoro, P.; Camacho, J.; Maciá-Fernández, G.; Gómez-Hernández, J.A.; López-Marín, V.J. A novel zero-trust network access control scheme based on the security profile of devices and users. *Comput. Netw.* **2022**, *212*, 109068. [CrossRef]
134. Ali, B.; Gregory, M.A.; Li, S.; Dib, O.A. Implementing zero trust security with dual fuzzy methodology for trust-aware authentication and task offloading in multi-access edge computing. *Comput. Netw.* **2024**, *241*, 110197. [CrossRef]
135. Zyoud, B.; Lutfi, S.L. The Role of Information Security Culture in Zero Trust Adoption: Insights from UAE Organizations. *IEEE Access* **2024**, *12*, 72420–72444. [CrossRef]
136. Khowaja, S.A.; Khuwaja, P.; Dev, K.; Singh, K.; Nkenyereye, L.; Kilper, D. ZETA: ZEro-Trust Attack Framework with Split Learning for Autonomous Vehicles in 6G Networks. In Proceedings of the 2024 IEEE Wireless Communications and Networking Conference (WCNC), Dubai, United Arab Emirates, 21–24 April 2024; pp. 1–6.
137. Meng, L.; Huang, D.; An, J.; Zhou, X.; Lin, F. A continuous authentication protocol without trust authority for zero trust architecture. *China Commun.* **2022**, *19*, 198–213. [CrossRef]
138. Xiao, S.; Ye, Y.; Kanwal, N.; Newe, T.; Lee, B. SoK: Context and risk aware access control for zero trust systems. *Secur. Commun. Netw.* **2022**, *2022*, 7026779. [CrossRef]
139. Cao, Y.; Pokhrel, S.R.; Zhu, Y.; Doss, R.; Li, G. Automation and orchestration of zero trust architecture: Potential solutions and challenges. *Mach. Intell. Res.* **2024**, *21*, 294–317. [CrossRef]
140. Liu, Z.; Li, X.; Mu, D. Data-Driven Zero Trust Key Algorithm. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 8659428. [CrossRef]
141. Raheman, F. From Standard Policy-Based Zero Trust to Absolute Zero Trust (AZT): A Quantum Leap to Q-Day Security. *J. Comput. Commun.* **2024**, *12*, 252–282. [CrossRef]
142. Ma, Z.; Chen, X.; Sun, T.; Wang, X.; Wu, Y.C.; Zhou, M. Blockchain-Based Zero-Trust Supply Chain Security Integrated with Deep Reinforcement Learning for Inventory Optimization. *Future Internet* **2024**, *16*, 163. [CrossRef]
143. Matiushin, I.; Korkhov, V. Continuous authentication methods for zero-trust cybersecurity architecture. In Proceedings of the International Conference on Computational Science and Its Applications, Athens, Greece, 3–6 July 2023; pp. 334–351.
144. Hosney, E.S.; Halim, I.T.A.; Yousef, A.H. An artificial intelligence approach for deploying ZTA. In Proceedings of the 2022 5th International Conference on Computing and Informatics (ICCI), Cairo, Egypt, 9–10 March 2022; pp. 343–350.
145. Singh, N.; Pal, S.; Leupers, R.; Merchant, F.; Rebeiro, C. PROMISE: A Programmable Hardware Monitor for Secure Execution in Zero Trust Networks. *IEEE Embed. Syst. Lett.* **2024**, *16*, 433–436. [CrossRef]
146. Hasan, S.; Amundson, I.; Hardin, D. Zero-trust design and assurance patterns for cyber-physical systems. *J. Syst. Archit.* **2024**, *155*, 103261. [CrossRef]
147. Liu, W.; Zhang, Z.; Qiao, X.; Li, Y.; Tan, Y.a.; Meng, W. A Software Integrity Authentication Protocol for Zero Trust Architecture. In Proceedings of the SIGCOMM Workshop on Zero Trust Architecture for Next Generation Communications, Sydney, Australia, 4 August 2024; pp. 1–6.
148. Sharma, S.; Singh, J.; Gupta, A.; Ali, F.; Khan, F.; Kwak, D. User Safety and Security in the Metaverse: A Critical Review. *IEEE Open J. Commun. Soc.* **2024**, *5*, 5467–5487. [CrossRef]
149. Munasinghe, S.; Piyarathna, N.; Wijerathne, E.; Jayasinghe, U.; Namal, S. Machine Learning Based Zero Trust Architecture for Secure Networking. In Proceedings of the 2023 IEEE 17th International Conference on Industrial and Information Systems (ICIIS), Peradeniya, Sri Lanka, 25–26 August 2023; pp. 1–6.
150. Spencer, M.; Pizio, D. The de-perimeterisation of information security: The Jericho Forum, zero trust, and narrativity. *Soc. Stud. Sci.* **2023**, *54*, 655–677. [CrossRef]
151. Shaik, M.; Gudala, L.; Sadhu, A.K.R. Leveraging Artificial Intelligence for Enhanced Identity and Access Management within Zero Trust Security Architectures: A Focus on User Behavior Analytics and Adaptive Authentication. *Aust. J. Mach. Learn. Res. Appl.* **2023**, *3*, 1–31.
152. Qazi, F.A. Study of zero trust architecture for applications and network security. In Proceedings of the 2022 IEEE 19th International Conference on Smart Communities: Improving Quality of Life Using ICT, IoT and AI (HONET), Marietta, GA, USA, 19–21 December 2022; pp. 111–116.

153. Katsis, C.; Cicala, F.; Thomsen, D.; Ringo, N.; Bertino, E. NEUTRON: A graph-based pipeline for zero-trust network architectures. In Proceedings of the Twelfth ACM Conference on Data and Application Security and Privacy, Baltimore, MA, USA, 24–27 April 2022; pp. 167–178.
154. Guo, J.; Xu, M. ZTESA-A Zero-Trust Endogenous Safety Architecture: Gain the endogenous safety benefit, avoid insider threats. In Proceedings of the International Symposium on Computer Applications and Information Systems (ISCAIS 2022), Shenzhen, China, 25–27 February 2022; Volume 12250, pp. 192–202.
155. Fang, L.; Wu, C.; Kang, Y.; Ou, W.; Zhou, D.; Ye, J. Zero-Trust-Based Protection Scheme for Users in Internet of Vehicles. *Secur. Commun. Netw.* **2022**, *2022*, 9896689. [CrossRef]
156. Neale, C.; Kennedy, I.; Price, B.; Yu, Y.; Nuseibeh, B. The case for zero trust digital forensics. *Forensic Sci. Int. Digit. Investig.* **2022**, *40*, 301352. [CrossRef]
157. Pigola, A.; de Souza Meirelles, F.; da Costa, P.R.; Porto, G.S. Trust in information security technology: An intellectual property analysis. *World Pat. Inf.* **2024**, *78*, 102281. [CrossRef]
158. Asad, M.; Otoum, S. Integrative Federated Learning and Zero-Trust Approach for Secure Wireless Communications. *IEEE Wirel. Commun.* **2024**, *31*, 14–20. [CrossRef]
159. Tang, F.; Ma, C.; Cheng, K. Privacy-preserving authentication scheme based on zero trust architecture. *Digit. Commun. Netw.* **2023**, *10*, 1211–1220. [CrossRef]
160. Li, P.; Ou, W.; Liang, H.; Han, W.; Zhang, Q.; Zeng, G. A zero trust and blockchain-based defense model for smart electric vehicle chargers. *J. Netw. Comput. Appl.* **2023**, *213*, 103599. [CrossRef]
161. Rivera, J.J.D.; Akbar, W.; Khan, T.A.; Muhammad, A.; Song, W.C. Secure enrollment token delivery mechanism for zero trust networks using blockchain. *IEICE Trans. Commun.* **2023**, *106*, 1293–1301. [CrossRef]
162. West, R.W.; Van der Merwe, J. dNextG: A Zero-Trust Decentralized Mobile Network User Plane. In Proceedings of the 19th ACM International Symposium on QoS and Security for Wireless and Mobile Networks, Montreal, QC, Canada, 30 October–3 November 2023; pp. 15–24.
163. Adahman, Z.; Malik, A.W.; Anwar, Z. An analysis of zero-trust architecture and its cost-effectiveness for organizational security. *Comput. Secur.* **2022**, *122*, 102911. [CrossRef]
164. Phiayura, P.; Teerakanok, S. A comprehensive framework for migrating to zero trust architecture. *IEEE Access* **2023**, *11*, 19487–19511. [CrossRef]
165. Singh, R.; Srivastav, G.; Kashyap, R.; Vats, S. Study on Zero-Trust Architecture, Application Areas & Challenges of 6G Technology in Future. In Proceedings of the 2023 International Conference on Disruptive Technologies (ICDT), Greater Noida, India, 11–12 May 2023; pp. 375–380.
166. Liu, Y.; Xing, X.; Tong, Z.; Lin, X.; Chen, J.; Guan, Z.; Wu, Q.; Susilo, W. Secure and scalable cross-domain data sharing in zero-trust cloud-edge-end environment based on sharding blockchain. *IEEE Trans. Dependable Secur. Comput.* **2023**, *21*, 2603–2618. [CrossRef]
167. Alagappan, A.; Venkatachary, S.K.; Andrews, L.J.B. Augmenting Zero Trust Network Architecture to enhance security in virtual power plants. *Energy Rep.* **2022**, *8*, 1309–1320. [CrossRef]
168. Gaber, M.G.; Ahmed, M.; Janicke, H. Malware detection with artificial intelligence: A systematic literature review. *ACM Comput. Surv.* **2024**, *56*, 1–33. [CrossRef]
169. Deng, G.; Liu, Y.; Mayoral-Vilches, V.; Liu, P.; Li, Y.; Xu, Y.; Zhang, T.; Liu, Y.; Pinzger, M.; Rass, S. {PentestGPT}: Evaluating and harnessing large language models for automated penetration testing. In Proceedings of the 33rd USENIX Security Symposium (USENIX Security 24), Philadelphia, PA, USA, 14–16 August 2024; pp. 847–864.
170. Masood, M.; Nawaz, M.; Malik, K.M.; Javed, A.; Irtaza, A.; Malik, H. Deepfakes generation and detection: State-of-the-art, open challenges, countermeasures, and way forward. *Appl. Intell.* **2023**, *53*, 3974–4026. [CrossRef]
171. Liang, C.; Du, H.; Sun, Y.; Niyato, D.; Kang, J.; Zhao, D.; Imran, M.A. Generative AI-driven semantic communication networks: Architecture, technologies and applications. *IEEE Trans. Cogn. Commun. Netw.* **2024**, *11*, 27–47. [CrossRef]
172. Wang, M.; Yang, N.; Forcade-Perkins, N.J.; Weng, N. Progen: Projection-based adversarial attack generation against network intrusion detection. *IEEE Trans. Inf. Forensics Secur.* **2024**, *19*, 5476–5491. [CrossRef]
173. Klasén, L.; Fock, N.; Forchheimer, R. The invisible evidence: Digital forensics as key to solving crimes in the digital age. *Forensic Sci. Int.* **2024**, *362*, 112133. [CrossRef]
174. Aboukadri, S.; Ouaddah, A.; Mezrioui, A. Machine learning in identity and access management systems: Survey and deep dive. *Comput. Secur.* **2024**, *139*, 103729. [CrossRef]
175. McIntosh, T.R.; Susnjak, T.; Liu, T.; Watters, P.; Xu, D.; Liu, D.; Nowrozy, R.; Halgamuge, M.N. From cobit to iso 42001: Evaluating cybersecurity frameworks for opportunities, risks, and regulatory compliance in commercializing large language models. *Comput. Secur.* **2024**, *144*, 103964. [CrossRef]

176. Fang, W.; Guan, X. Research on iOS remote security access technology based on zero trust. In Proceedings of the 2022 IEEE 6th Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing, China, 4–6 March 2022; Volume 6, pp. 238–241.
177. Du, X.; Hargreaves, C.; Sheppard, J.; Scanlon, M. TraceGen: User activity emulation for digital forensic test image generation. *Forensic Sci. Int. Digit. Investig.* **2021**, *38*, 301133. [CrossRef]
178. Begou, N.; Vinoy, J.; Duda, A.; Korczyński, M. Exploring the dark side of ai: Advanced phishing attack design and deployment using chatgpt. In Proceedings of the 2023 IEEE Conference on Communications and Network Security (CNS), Orlando, FL, USA, 2–5 October 2023; pp. 1–6.
179. Ayoobi, N.; Shahriar, S.; Mukherjee, A. The looming threat of fake and llm-generated linkedin profiles: Challenges and opportunities for detection and prevention. In Proceedings of the 34th ACM Conference on Hypertext and Social Media, Rome, Italy, 4–8 September 2023; pp. 1–10.
180. Frankovits, G.; Mirsky, Y. Discussion paper: The threat of real time deepfakes. In Proceedings of the 2nd Workshop on Security Implications of Deepfakes and Cheapfakes, Melbourne, VIC, Australia, 10–14 July 2023; pp. 20–23.
181. Schmitt, M.; Flechais, I. Digital Deception: Generative artificial intelligence in social engineering and phishing. *Artif. Intell. Rev.* **2024**, *57*, 1–23. [CrossRef]
182. Chen, Y.; Cui, M.; Wang, D.; Cao, Y.; Yang, P.; Jiang, B.; Lu, Z.; Liu, B. A survey of large language models for cyber threat detection. *Comput. Secur.* **2024**, *145*, 104016. [CrossRef]
183. Zhao, C.; Du, H.; Niyato, D.; Kang, J.; Xiong, Z.; Kim, D.I.; Shen, X.; Letaief, K.B. Generative AI for secure physical layer communications: A survey. *IEEE Trans. Cogn. Commun. Netw.* **2024**, *11*, 3–26. [CrossRef]
184. Mirsky, Y.; Demontis, A.; Kotak, J.; Shankar, R.; Gelei, D.; Yang, L.; Zhang, X.; Pintor, M.; Lee, W.; Elovici, Y.; et al. The threat of offensive ai to organizations. *Comput. Secur.* **2023**, *124*, 103006. [CrossRef]