Review Article

Zero-Trust Architecture (ZTA): Designing an AI-Powered Cloud Security Framework for LLMs' Black Box Problems

Bibhu Dash[*]

School of Computer and Information Sciences, University of the Cumberlands, USA

Abstract

Businesses are becoming more interested in developing and testing Large Language Models (LLMs) in their own settings to support decision-making and growth as a result of the rapid emergence of AI and cloud computing. Here's the dilemma, though: to what extent do you believe these models and the data they were trained on? We don't know the feature list of an LLM, which presents the first obstacle when discussing trust and the reasons why there should be zero trust. Although it may seem a bit extreme, this is accurate for two reasons. When it comes to GenAI models nowadays, the more multimodal and more capabilities they have, the better. This way of thinking is great for exploring and confirming if GenAI can address a business problem, but it's a surefire way to run into trouble when attempting to put things into production in an organizational setting. An enterprise cybersecurity architecture known as a zero-trust architecture (ZTA) is built on the ideas of zero trust and is intended to stop data breaches, enhance privacy, and restrict internal lateral movement. This article discusses ZTA, its logical aspects, probable deployment scenarios, AI rules, threats and limitations in order to provide a detailed understanding of why enterprises must adapt a ZTA framework in a cloud-based environment for AI model deployment.

Introduction

According to the Zero Trust security architecture, before obtaining or keeping access to apps and data, all users - both inside and outside the company's network must be validated, given permission, and regularly assessed for security configuration and posture [1]. Since resources and employees can be situated anywhere, networks can be local, cloud-based, or a combination of both, Zero Trust assumes that there is no such thing as a normal network edge. In today's digital age, zero trust is a process that protects data and infrastructure. It addresses modern business issues including ransomware threats, hybrid cloud environments, and security for remote workers in a new way. Although many providers have endeavored to delineate Zero Trust in their own manner, certain established organizations' guidelines might aid you in harmonizing Zero Trust with your establishment.

In response to the increasing number of high-profile cyber breaches, the US government issued an executive order in May 2021 requiring U.S. Federal Agencies to comply with National Institute of Standards and Technology (NIST) 800-207 as an essential step for Zero Trust implementation [2,3]. Numerous commercial organizations, government agencies, and vendors have provided considerable validation and input on the standard, making it the de facto norm for private firms as well.

Zero Trust in cloud as per NIST guidelines [2], follow these 3 characteristics:

a) Ongoing verification: Make sure you always have access to all resources.
b) Restrict 'blast radius': Reduce the damage if an insider or external breach occurs.
c) Automate the gathering and reaction of context: For the most correct action, consider behavioral data and obtain context from the complete IT stack (identity, endpoint, workload, etc.).
d) Real-time data analytics and monitoring: Validate and take real time action on the spot with analytics to safeguard the IT assemble.

Why ZTA Needed for Cloud?

As cloud is everywhere in modern organizations, through cloud-based Zero Trust Architecture (ZTA) implementation, enterprises may take a more proactive and detailed approach to security. Identity-based access control, continuous authentication, and micro-segmentation are a few examples of Zero Trust principles that assist enterprises gain better agility and flexibility while reducing the dangers associated with cloud computing (Figure 1) [4]. In cloud contexts, Zero Trust Architecture is becoming more and more crucial for several reasons:

a) Distributed Nature: Resources are accessed from a variety of places and devices in cloud settings, which frequently span numerous regions and data centers. Because of this distributed nature, traditional security methods focused on the network perimeter become less effective as the boundary grows more porous [3,4].
b) Dynamic Workload Settings: Workloads are spun up and down in response to demand in cloud settings, which are quite dynamic. This degree of dynamism defies the capabilities of traditional security techniques, which are based on static network boundaries [5].
c) Expanded Attack Surface: As cloud services are adopted and remote work becomes more common, enterprises now confront a larger attack surface [5]. Attackers now have greater chances to take advantage of weaknesses and access cloud resources without authorization.
d) Changing Threat Environment: Attackers' strategies are getting more complex, and cyberthreats are always changing. To counter these changing threats, traditional security methods that are predicated on static trust assumptions are inadequate [6].

e) Data Privacy and Compliance: Organizations are under growing pressure to guarantee the security and privacy of sensitive data due to laws like the General Data Protection Regulation (GDPR) and California Consumer Privacy Act (CCPA). Least privilege access and data encryption are two examples of zero trust concepts that aid firms in meeting these compliance obligations [1,7].

Figure 1: Enterprise with remote and onsite employee behaviors [3].

Role of AI in ZTA

By offering cutting-edge capabilities for threat detection, access control, and incident response, Artificial Intelligence (AI) technologies can greatly increase the efficacy of Zero Trust Architecture. This will help organizations become more resilient against cyber threats in the ever-changing and dynamic threat landscape of today. AI can dramatically improve Zero Trust Architecture by allowing for better threat detection, access management, and anomaly identification. Here's how AI can help you implement Zero Trust principles:

a) **Behavioral Analytics:** To create a baseline of typical activity, AI-powered behavioral analytics can examine user and entity behavior throughout the network. Organizations can identify and react to unusual activity in real-time by flagging any departures from this baseline as possible security risks [1].

b) **Continuous Authentication:** AI can facilitate continuous authentication by continuously confirming the identity of users gaining access to network resources by evaluating user behavior patterns, including typing speed, mouse movements, and biometric data. This aids in preventing unwanted access even following the first authentication [2].

c) **Threat Intelligence:** To recognize new threats and take preemptive measures to counter them, artificial intelligence (AI) algorithms can examine enormous volumes of threat intelligence data from a variety of sources, such as security feeds, dark web monitoring, and previous attack data. Organizations may improve their security posture and remain ahead of changing threats by incorporating AI-driven threat intelligence into the Zero Trust architecture [8].

d) **Access Control and Privilege Management:** Privilege management and access control are made possible by AI-based systems that may dynamically modify access privileges in response to risk assessments, contextual variables, and user behavior [2,3]. Organizations may minimize the risk of privilege misuse or credential compromise while guaranteeing that users always have the right amount of access to resources by using AI to automate access control decisions.

e) **Network Segmentation and Micro-Segmentation:** Using user roles, application dependencies, and security policies as guidelines, AI-powered network segmentation systems may autonomously divide a network. Thus, the attack surface is decreased, and the consequences of security breaches are minimized. This also helps enterprises to impose strict access rules and isolate vital assets from possible threats [5].

f) **Automated Threat Response:** By automatically coordinating security alerts, ranking threats, and planning response actions, AI-driven security orchestration and automation solutions can expedite incident response procedures [3]. This reduces the possible impact on business operations by assisting enterprises in identifying and mitigating security events more quickly.
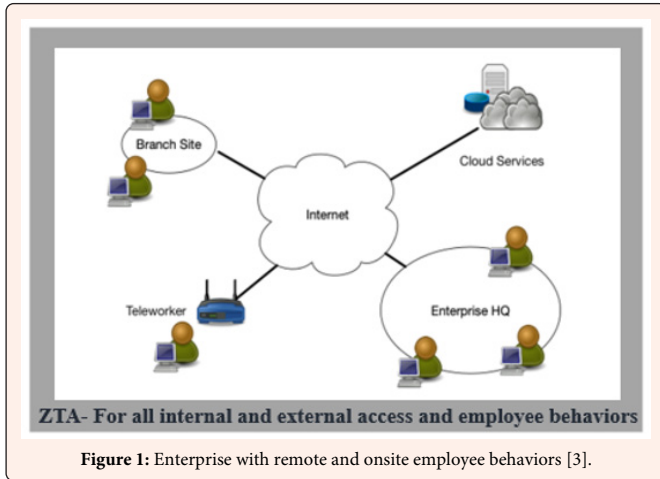
LLMs Black Box Problems

Research on "black box" problems in LLMs is ongoing, with the goal of enhancing the robustness, transparency, and interpretability of the models [9]. Nevertheless, comprehensive answers to these problems are still elusive, and properly implementing LLMs necessitates giving these problems considerable thought. Occasionally referred to as "black box" issues, large language models (LLMs) such as GPT (Generative Pre-trained Transformer) models might display certain behaviors. This is the reason why:

a) **Complexity:** With millions or perhaps billions of parameters and data points, LLMs are extremely complicated models. It can be difficult to understand exactly how they produce outputs or make judgments because of their intricacy [9,10].

b) **Lack of Transparency:** Although efforts have been made to increase LLMs' transparency, such as by highlighting the most important portions of the input with attention techniques, these models' internal workings can still be opaque [10,11]. Because of this lack of transparency, it may be challenging to comprehend the reasoning behind a certain decision, apprehension or outcome.

c) **Limited Explanation:** Although various tools, like attention visualization or probing techniques, are available to explain LLMs, these approaches might not fully capture the model's reasoning processes [11]. As a result, even if they might shed light on model behavior, they don't provide full explanation.

d) **Fairness and Bias Problems:** LLMs may unintentionally pick up on and reinforce biases found in the training data. It can be difficult to recognize and address these biases, especially when the model's decision-making process is opaque [10,12,13].

e) **Robustness and Vulnerability:** It has been proved that LLMs are vulnerable to adversarial attacks, in which even little changes to the input can produce noticeably different results. Research is still being done to decide why these vulnerabilities exist and how to address them [14-17].

How Zero Trust Works

In traditional security model, users and endpoints that were inside the company's network perimeter were taken for granted under the conventional paradigm and externals use Virtual Private Networks (VPNs) [18]. This made the company vulnerable to rogue credentials and hostile internal actors, and once unauthorized users were inside the network, it unintentionally gave them broad access. For businesses with several cloud environments, multiple linked systems, and a desire for more control over individual cloud access as well as cloud-based services and apps, a Zero Trust architecture is perfect [18,19]. Compared to traditional network security, which employed the "trust but verify" approach, zero trust stands for a significant change [20,21]. The implementation of this framework combines advanced technologies such as risk-based multi-factor authentication, identity protection, next-generation endpoint security, and robust cloud workload technology to verify a user or system's identity, consider access at that time, and keep system security [20]. Zero Trust also demands data encryption, email security, and asset and endpoint cleanliness verification before connecting to applications [21]. Under Zero Trust architecture, organizations must thus constantly monitor and verify that an individual and their device (real-time validation) have the required privileges and attributes [22]. Prior to authorizing the transaction, it also calls for policy enforcement that considers compliance requirements, user and device risk, and other factors. It requires that the company be able to apply controls on what and where its privileged accounts connect, as well as be aware of all of its service and account details. Because threats and user traits are dynamic, one-time validation is insufficient [23].

To enhance algorithmic AI/ML model training for ultra-accurate policy response, analytics must be connected to billions of events, extensive corporate telemetry, and threat intelligence. Conducting a comprehensive evaluation of their IT infrastructure and possible avenues for attack can help organizations prevent assaults and lessen the impact of security breaches. Segmentation according to device kind, identity, or group functions

Citation: Dash B (2024) Zero-Trust Architecture (ZTA): Designing an AI-Powered Cloud Security Framework for LLMs' Black Box Problems. Current Trends in Eng Sci. 4: 1058

Page 2/5

may be necessary for this. For instance, dubious protocols like RDP or RPC (Remote Procedure Call) to the domain controller ought to be restricted to specific credentials or constantly questioned [22,23]. Over 80% of all attacks entail the exploitation or abuse of credentials within the network [24]. Increased password security, account integrity, following corporate policies, and avoiding high-risk shadow IT services are all made possible with the aid of ZTA (Figure 2).
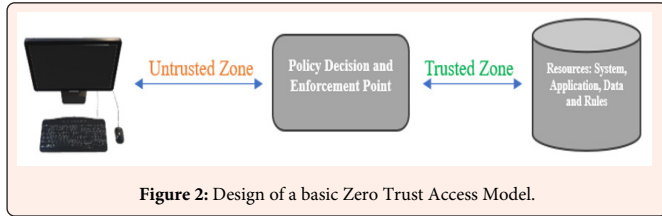

Figure 2: Design of a basic Zero Trust Access Model.

Designing an AI-Powered Zero Trust Framework

Working with AI models, particularly LLMs, needs a large amount of data, and we have no control over what the models use internally. All IT engineers and data scientists are questioning whether the models internally store the data, delete it later, or share it. This is purely a black box to any IT specialist. Hence, creating an AI-driven Zero Trust Framework for cloud settings to work with LLMs necessitates a thorough strategy that considers the particular difficulties in overseeing security in data sharing across various cloud platforms [2,25]. Organizations may build a strong Zero Trust Framework for single or multi-cloud environments that uses AI to improve security, visibility, and compliance across their cloud deployments by incorporating AI technology into these essential elements (Figure 3).

Unified Identity and Access Management (UIAM)

a) **Centralized Identity Federation:** Unified Identity and Access Management (UIAM) install a centralized identity federation system to serve as a single source of truth for user IDs and access restrictions in all cloud environments. This system should interface with numerous cloud identity providers (IdPs) and directories [25].

b) **AI-driven Access Policies:** Access policies driven by artificial intelligence can be created and enforced dynamically in cloud or multi-cloud systems by using AI algorithms to analyze user behavior, contextual factors, and compliance requirements [26].

Orchestration of Network Security

a) **AI-driven Network Segmentation:** Implement AI-driven network segmentation systems to automatically divide traffic and apply security rules according to user roles, workload characteristics, and application dependencies in a variety of cloud environments [27].

b) **Activity-based Threat Detection:** In real-time, across multi-cloud settings, monitor network traffic and spot unusual activity suggestive of cyber threats, such lateral movement or data exfiltration, by utilizing AI-driven threat detection algorithms [22].

Endpoint protection and device administration

a) **AI-driven Endpoint Protection:** Use machine learning algorithms to detect and respond to emerging threats by deploying AI-powered endpoint protection platforms that offer centralized visibility and control over endpoints and devices spread across multi-cloud environments [3].

b) **Device Compliance Automation:** To ensure a consistent security posture across various cloud environments, use AI-driven compliance automation solutions to evaluate and enforce device compliance with security policies and configurations [2].

Data Security and Encryption

a) **AI-driven Data Classification:** Using metadata properties, content, and context, you may use AI-driven data classification technologies to automatically detect and categorize sensitive data that is stored across several cloud environments [24].

b) **Dynamic Data Encryption:** To safeguard sensitive data while it's in use, in transit, and at rest across various cloud platforms, as well as to guarantee confidentiality and integrity, employ AI algorithms for key management and dynamic data encryption [22,24].

Threat Intelligence and Incident Response

a) **Cloud Threat Intelligence Integration:** Integration of AI-driven threat intelligence feeds from various sources can offer thorough insight into cyberthreats and vulnerabilities impacting multi-cloud settings, facilitating proactive threat hunting and incident response. This is known as multi-cloud threat intelligence integration [2,3].

b) **Cross-Cloud Incident Orchestration:** To expedite response times and lessen the impact of security incidents, deploy AI-driven security orchestration and automation platforms that can coordinate incident response actions across various cloud environments, such as alert triaging, threat containment, and remediation [3].

Continuous Monitoring and Compliance Assurance

a) **AI-driven Security Analytics:** Use AI-driven security analytics tools to monitor and detect threats continuously by analyzing security logs, events, and telemetry data from various cloud platforms to spot unusual patterns and signs of compromise [24].

b) **Cloud Compliance Automation:** To guarantee adherence to legal requirements and industry standards, use AI-driven compliance automation solutions to automate compliance evaluation and enforcement across several cloud environments [27].
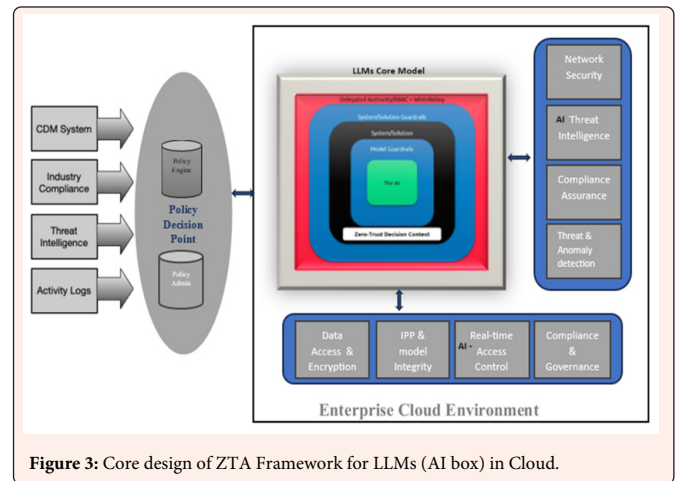

Figure 3: Core design of ZTA Framework for LLMs (AI box) in Cloud.

In Figure 3, the core design of ZTA logic uses separate control plane to communicate while application data communicate on data plane to the AI models shown in a small green box at the center. As the core strategy for data accessibility and output are not clearly said, the policy decision point (PDP) with help of policy engine and policy admin, takes the charge of final decision before passing data to the LLM models through cloud setting as shown above [28]. Along with PDP, the system administrator applies the security rules on each input data or access points to models by ZTA deployed just before the model guardrails.

Special Security around LLMs in Cloud

With popularity around ChatGPT, organizations are excited to build and test their own inhouse GPT models [17]. But considering LLMs black box problems, it needs a complete strategy to secure Large Language Models (LLMs) in a cloud context, addressing issues with data privacy, model integrity, access control, and compliance, among other security problems (Figure 4). When building Zero Trust Architecture in cloud, along with above points, special security around LLMs in a cloud or multi-cloud environment, keep the following points in view:

a) **Data Encryption and Privacy:** To prevent unwanted access or interception of sensitive information, encrypt LLM data both in transit and at rest [28]. To anonymize sensitive data used to train LLMs, use data masking or tokenization techniques, particularly if the training data holds personally identifiable information (PII). Obtain user consent for any data processing operations involving LLMs and implement necessary data privacy controls to ensure compliance with data protection legislation (e.g., GDPR, CCPA) [27,28].

b) **Model Integrity and Intellectual Property Protection (IPP):** Implement strong access controls and encryption techniques to safeguard LLM models against theft of intellectual property, tampering, and illegal access [29]. Make sure to guarantee the integrity of deployed LLM instances, make use of secure model deployment and hosting platforms that include features like code signing, containerization, and secure bootstrapping. To enforce licensing and usage limits for LLM models, particularly in commercial applications where intellectual property protection is critical, implement digital rights management (DRM) technologies.

c) **Access Control and Authentication:** Give focus to role-based access control (RBAC) policies and fine-grained access controls to limit access to LLM resources according to the least privilege, user roles, and privileges principles. To confirm the identity of users and applications accessing LLM APIs or administrative interfaces, employ robust authentication techniques (e.g., OAuth, OpenID Connect) and multi-factor authentication (MFA) [30,31]. Using centralized logging and monitoring systems, keep an eye on and audit user access to LLM resources to spot and handle any unwanted access attempts or questionable activity.

d) **Threat and Anomaly Monitoring:** Use artificial intelligence (AI)-driven threat detection and anomaly monitoring tools to find malicious activity, illegal data access, or unusual patterns of behavior linked to LLMs. Keep an eye out for indications of compromise or security events in system logs, network traffic, and API usage patterns. Integrate anomaly detection alerts with security incident response workflows to mitigate security issues promptly. Use security analytics tools and threat intelligence feeds to keep up with new attacks and vulnerabilities that could compromise LLM deployments in multi-cloud settings [31].

e) **Apply Compliance and Governance:** Assure adherence to cloud provider security requirements (i.e. AWS Well-Architected Framework or Azure Security Benchmark -ASB) and industry-specific laws such as HIPAA and PCI DSS those are pertinent to LLM implementations. To evaluate and reduce security risks related to LLM usage, put in place risk management procedures and governance frameworks, including as data governance, model governance, and third-party risk management [2,3,31]. Conduct routine compliance audits, penetration tests, and security assessments to confirm the efficacy of security measures and pinpoint opportunities for enhancing the protection of LLMs in multi-cloud settings [32].
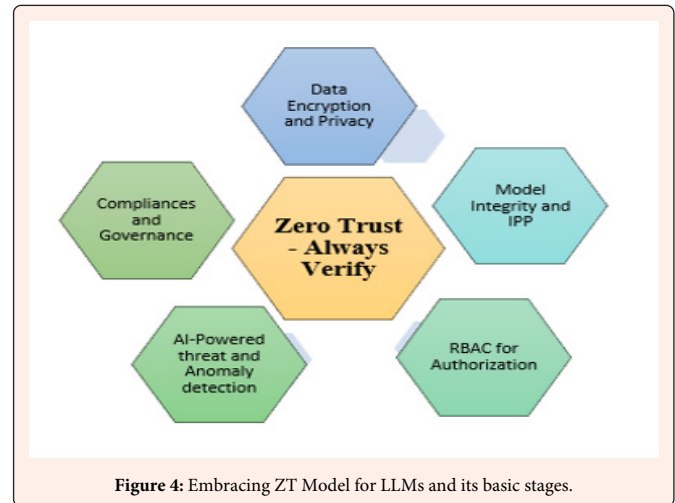

Figure 4: Embracing ZT Model for LLMs and its basic stages.

Limitations of Building a ZTA on Cloud

Building a 100% Zero Trust Framework is too difficult due to its growing design complexity and integration of several third-party integrations in modern enterprises. ZTA differs depending on the technologies, techniques or rules, and approaches used by businesses to achieve it [33]. ZTA implementation is particularly difficult for many organizations since they are still reliant on legacy tools and technology. limitations on resources can make it difficult to scale ZTA solutions to meet expanding cloud deployments, rising workloads, and rising user volumes. In dispersed cloud systems, it needs speed optimization, low latency, and high availability maintenance, that puts a lot of pressure on ZTA design and performance. While some businesses largely rely on ZTA solutions without conducting adequate testing or comprehending its design patterns, others claim that their ZTA model does not support their internal access and storage management policies [34]. Too much reliance and dependability are dangerous for organizational data and privacy security since the assets used to store and process this Zero Trust information frequently lack an open standard for how to interact and exchange information.

In order to achieve an effective security posture in cloud environments, organizations must carefully consider and address the limitations and challenges related to implementation, integration, scalability, cost, vendor lock-in, regulatory compliance, and user experience - even though deploying a ZTA in the cloud offers significant security benefits [33].

Conclusion

In summary, developing a cloud based Zero-Trust Architecture framework driven by AI provides a solid and adaptable strategy for improving security posture in contemporary digital settings. Organizations can attain real-time threat detection, adaptive response capabilities, and granular access control across various cloud platforms by utilizing artificial intelligence (AI) technologies like machine learning, behavioral analytics, and threat intelligence integration. But putting into practice a ZTA framework driven by AI necessitates giving careful thought to a number of variables, including user

experience impact, regulatory compliance, scalability, and integration complexity. For enterprises looking to safeguard vital assets, lessen cyber threats, and guarantee data privacy in cloud settings, adopting AI-powered ZTA frameworks is essential despite these obstacles due to the advantages of increased security, agility, and resilience.

Acknowledgement

References

1. Stafford VA (2020) Zero trust architecture. NIST special publication.

2. Shastri V (2023) What is ZTNA? Zero Trust Network Access – Crowdstrike.

3. Kerman A, Borchert O, Rose S, Tan A (2020) Implementing a zero-trust architecture. National Institute of Standards and Technology p. 17.

4. Teerakanok S, Uehara T, Inomata A (2021) Migrating to zero trust architecture: Reviews and challenges. Security and Communication Networks p. 1-10.

5. Shastri V (2023) What is ZTNA? Zero Trust Network Access - Crowdstrike.

6. Loftus M, Vezina A, Doten R, Mashatan A (2023) The Arrival of Zero Trust: What Does it Mean? Communications of the ACM 66(2): 56-62.

7. Kawalkar SA, Bhoyar DB (2024) Design of an Efficient Cloud Security Model through Federated Learning, Blockchain, AI-Driven Policies, and Zero Trust Frameworks. International Journal of Intelligent Systems and Applications in Engineering 12(10s): 378-388.

8. Grassi L, Recchiuto CT, Sgorbissa A (2023) Sustainable cloud services for verbal interaction with embodied agents. Intelligent Service Robotics 16(5): 599-618.

9. Brożek B, Furman M, Jakubiec M, Kucharzyk B (2023) The black box problem revisited. Real and imaginary challenges for automated legal decision making. Artificial Intelligence and Law pp. 1-14.

10. Madsen T (2024) Zero-trust–An Introduction. CRC Press.

11. Kumar A, Singh S, Murty SV, Ragupathy S (2024) The Ethics of Interaction: Mitigating Security Threats in LLMs.

12. Wrana M, Barradas D, Asokan N (2024) The Spectre of Surveillance and Censorship in Future Internet Architectures.

13. Tsai YHH, Talbott W, Zhang J (2024) Efficient Non-Parametric Uncertainty Quantification for Black-Box Large Language Models and Decision Planning.

14. Michaud EJ, Liao I, Lad V, Liu Z, Mudide A, et al. (2024) Opening the AI black box: program synthesis via mechanistic interpretability.

15. Wang Y, Ma X, Chen W (2023) Augmenting black-box llms with medical textbooks for clinical question answering.

16. Cheng J, Liu X, Zheng K, Ke P, Wang H, et al. (2023) Black-box prompt optimization: Aligning large language models without model training.

17. Sharma P, Dash B (2023) Impact of big data analytics and ChatGPT on cybersecurity. In 2023 4th International Conference on Computing and Communication Systems (I3CS) IEEE pp. 1-6.

18. Wu X, Wu SH, Wu J, Feng L, Tan KC (2024) Evolutionary Computation in the Era of Large Language Model: Survey and Roadmap.

19. Hassija V, Chamola V, Mahapatra A, Singal A, Goel D, et al. (2024) Interpreting black-box models: a review on explainable artificial intelligence. Cognitive Computation 16(1): 45-74.

20. Saleem M, Warsi MR, Islam S (2023) Secure information processing for multimedia forensics using zero-trust security model for large scale data analytics in SaaS cloud computing environment. Journal of Information Security and Applications 72: 103389.

21. Seaman J (2023) Zero Trust Security Strategies and Guideline. In Digital Transformation in Policing: The Promise, Perils and Solutions. Cham: Springer International Publishing, pp. 149-168.

22. Chauhan M, Shiaeles S (2023) An analysis of cloud security frameworks, problems and proposed solutions. Network 3(3): 422-450.

23. Pero V, Ekman L (2023) Implementing a Zero Trust Environmentfor an Existing On-premises Cloud Solution.

24. Dash B, Ullah S (2024) Quantum-safe: Cybersecurity in the age of Quantum-Powered AI. World Journal of Advanced Research and Reviews 21(1): 1555-1563.

25. Fernandez EB, Brazhuk A (2024) A critical analysis of Zero Trust Architecture (ZTA). Computer Standards & Interfaces.

26. Rodigari S (2023) Performance Analysis of Zero Trust in Cloud Native Systems.

27. Salminen H (2023) Zero Trust: The Magic Bullet or Devil's Advocate? In European Conference on Cyber Warfare and Security 22(1): 678-686.

28. Morrow T (2023) Best Practices and Results from Fall 2022 SEI Zero Trust Industry Day.

29. Kujo J (2023) Implementing Zero Trust Architecture for Identities and Endpoints with Microsoft tools.

30. Gai K, She Y, Zhu L, Choo KKR, Wan Z (2023) A blockchain-based access control scheme for zero trust cross-organizational data sharing. ACM Transactions on Internet Technology 23(3): 1-25.

31. Gao S, Gao AK (2023) On the Origin of LLMs: An Evolutionary Tree and Graph for 15,821 Large Language Models.

32. Al Shehhi F, Otoum S (2023) On the Feasibility of Zero-Trust Architecture in Assuring Security in Metaverse. In 2023 International Conference on Intelligent Metaverse Technologies & Applications (iMETA). IEEE pp. 1-8.

33. Syed NF, Shah SW, Shaghaghi A, Anwar A, Baig Z, et al. (2022) Zero trust architecture (zta): A comprehensive survey. IEEE Access 10: 57143-57179.

34. Feng X, Hu S (2023) Cyber-Physical Zero Trust Architecture for Industrial Cyber-Physical Systems. IEEE Transactions on Industrial Cyber-Physical Systems 1: 394-405.

Citation: Dash B (2024) Zero-Trust Architecture (ZTA): Designing an AI-Powered Cloud Security Framework for LLMs' Black Box Problems. Current Trends in Eng Sci. 4: 1058

Page 5/5