laboratories

MDPI

Communication

Adapting Zero Trust to University Laboratory Safety Governance in the Artificial Intelligence Era

Kaixi Jiang

Office of Laboratory Safety and Equipment Management, Beijing Normal University, Zhuhai 519087, China; jiangkaixi@bnu.edu.cn

Abstract

In the artificial intelligence (AI) era, the structure of safety hazards in university laboratories is changing, with unconventional hazards making up a growing share. As AI evolves from an auxiliary tool into an autonomous research collaborator, interdisciplinary experiments are becoming faster and more intense, producing hazards that are more dynamic, less predictable, and harder to assess and prevent. Most laboratory safety frameworks, however, still rely on static rules built for known risks, and these are poorly suited to such emerging threats. A shift from static rule-based control toward dynamic risk management is therefore needed. Adopting a conceptual mapping approach rather than an empirical study, this paper examines how Zero Trust—a cybersecurity framework based on continuous trust evaluation—could be adapted to university laboratory safety. It defines three adapted attributes (universal distrust with dynamic verification, least privilege with dynamic adjustment, and continuous monitoring with real-time early warning), identifies the main implementation barriers, and suggests targeted responses.

Keywords: AI era; Zero Trust; laboratory safety; safety governance; dynamic risk control

1. Introduction

The rapid advancement and deep integration of AI technologies in university research are profoundly reshaping higher education research paradigms while improving research efficiency and innovation capacity. AI is no longer merely an auxiliary tool but has developed into an independent research partner, enabling closer and deeper human–machine collaboration [1]. This transformation has made interdisciplinary cooperation and its supporting experimental activities the mainstream research mode, and simultaneously given rise to new types of unconventional laboratory safety hazards marked by dynamism, unpredictability, difficulty in quantitative assessment and limited predictability.

Against this backdrop, university laboratory safety management is facing a complex situation where conventional and unconventional hazards coexist and compound one another. In this paper, conventional safety hazards are defined as risks that can be effectively identified, predicted and controlled through mature technologies and methodologies within the framework of conventional laboratory safety management systems and regulations [2]. By contrast, unconventional safety hazards are defined as risks that cannot be effectively identified, predicted or regulated by existing technologies and methods under conventional management systems and regulatory frameworks. The boundary between the two is whether a hazard lies within the predictive and controllable range of established, rule-based safety arrangements: once it can be identified, forecast and controlled with mature methods, it is treated as conventional; hazards driven by novel, poorly understood

Check for updates

Academic Editor: Gassan Hodaifa

Received: 20 August 2026
Revised: 22 September 2026
Accepted: 29 September 2026
Published: 8 October 2026

Copyright: © 2026 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

or rapidly evolving conditions that escape such methods are treated as unconventional. Different from conventional hazards, unconventional risks mainly arise from unknown factors and uncertainties inherent to interdisciplinary integration and cutting-edge exploratory experiments, thus frequently falling outside the scope of existing safety codes. Although no systematic incident statistics are yet available, on the basis of practical experience we expect—with a reasonable degree of probability rather than as an established empirical fact—that the proportion of unconventional hazards among all laboratory safety risks will continue to rise. We do not claim that existing management frameworks have failed systematically; our concern is that, as unconventional hazards increase in number and complexity, static, rule-based frameworks designed for known risks will find it progressively harder to identify, assess and provide timely support for these emerging threats. This evolutionary trend constitutes the logical starting point for constructing a new laboratory safety management system.

The governance of conventional safety hazards mainly relies on established safety management systems and regulatory frameworks to govern the safety conditions of core elements including laboratory personnel, mechanical equipment, experimental supplies, experimental methods and environmental conditions. However, this static rule-based management model presents notable limitations. For example, in terms of trust mechanisms, the model relies on voluntary compliance within the university-college-lab three-level responsibility system, often replacing dynamic full-process supervision with static institutional trust. From the perspective of risk assessment, standardized static regulatory frameworks fail to adapt to the dynamic evolution of risks, easily leading to risk misjudgment and ineffective prevention and control. Regarding laboratory access permission management, insufficient dynamic review and adjustment capacity, coupled with delayed authorization feedback, often results in two imbalanced governance extremes of over-restriction or inadequate supervision. In short, the existing system performs adequately for recurring, well-understood risks, but its core weaknesses—a static trust baseline, rule-driven rather than risk-driven control, and delayed feedback—are precisely the points most exposed by unconventional hazards.

Although the static rule-based management model remains largely applicable and effective for controlling conventional safety hazards in university laboratories to date, it cannot cope with unconventional safety hazards. To address this dilemma, it is urgent to establish a new overarching management framework to guide laboratory safety governance. Centered on the core principle of "never trust, always verify", Zero Trust has emerged as a highly innovative and practically valuable paradigm in the field of cybersecurity [3]. This paradigm is highly compatible with the new safety governance landscape of university laboratories in the AI era, and holds substantial potential for cross-domain application in laboratory safety administration.

It should also be clarified that this paper is a conceptual and perspective contribution rather than an empirical study. Its approach is one of conceptual mapping: starting from the Zero Trust paradigm in cybersecurity, it examines how the paradigm could be translated, adapted and provisionally organized for the laboratory safety domain, drawing on the author's long-standing practical experience in university laboratory safety administration. No new incident statistics, questionnaire data or controlled experiments are reported; the aim is to frame the problem, propose an adapted conceptual framework, and identify implementation priorities, challenges and possible strategies, with quantitative validation left to future work.

2. Adaptability of Zero Trust to University Laboratory Safety Management

Before proceeding, two terms are clarified. By safety we mean the set of measures and states intended to prevent harm to personnel, property and the environment in the laboratory, to detect deviations early, and to contain their consequences; it matters because a single serious incident may cause injury, loss of life, and the disruption of research. By governance we mean the broader system of rules, responsibilities, oversight and feedback through which safety risks are continuously managed—not a single inspection or a fixed set of regulations, but a cyclic process in which standards are set, compliance is verified, deviations are corrected, and arrangements are adjusted over time.

2.1. Ideological Connotation of Zero Trust

Conventional security philosophy is rooted in perimeter defense logic, dividing cyberspace into trusted intranets and untrusted external networks. Nevertheless, with the growing complexity of network environments and accelerated cloud migration, various dynamic, hidden and sophisticated security threats have emerged within previously trusted intranets, rendering conventional network security perimeters ineffective. In this context, cybersecurity researcher John Kindervag first proposed the Zero Trust concept and its corresponding architecture in 2010 [4], to challenge the static reliance of conventional cybersecurity on physical perimeter defense. Google designed and deployed an internal Zero Trust security architecture in 2014 [5]. In 2020, the U.S. National Institute of Standards and Technology (NIST) published formal standards, implementation specifications and real-world application cases for Zero Trust architectures [6]. At present, Zero Trust has been gradually extended from cybersecurity to various industrial scenarios, demonstrating its cross-scenario transferability and robust adaptability.

It is worth emphasizing that cybersecurity is not a single boundary but a stack of interdependent layers—physical and access control, network, host and system, application, data, and the identity layer that spans them. Perimeter defense protected only the outer boundary and assumed everything inside was trustworthy; Zero Trust by contrast operates across the identity, device, application and data layers, verifying every access request on the basis of identity, state and context rather than location. This layered, internal-verification logic is precisely what laboratory safety lacks, and is why the cross-domain transfer is promising rather than merely metaphorical.

The core of Zero Trust lies in reconstructing trust mechanisms: it operates on the principle that no access subject—including users, devices, systems, and applications—should be inherently trusted [4]. Instead, Zero Trust architecture conducts fine-grained multi-dimensional risk assessment for every access and operation request, encompassing identity, operational necessity, security state and historical behavior [4–6]. This new defensive principle characterized by proactive prevention, continuous verification and dynamic adjustment provides a novel perspective for enhancing the governance efficacy of both conventional and unconventional hazards in university laboratories.

2.2. Adaptability of Cross-Domain Integration of Zero Trust

Current university laboratory safety management is predominantly grounded in static institutional constraints, which struggle to address complex scenarios with overlapping and interacting conventional and unconventional hazards. While static regulations still remain a foundational regulatory mechanism in managing conventional hazards, they face long-standing governance bottlenecks and efficiency ceilings, as safety violations remain persistent and recurrent. For unconventional hazards, their suddenness and uncertainty render the conventional static rule-based restriction model largely ineffective.

Addressing the key challenge of recurrent violations under the conventional hazard management regulations [7], the Zero Trust principle of "never trust, always verify" can shift governance logic from one-time compliance audits to full-process dynamic verification. Real-time inspection of operational compliance can effectively reduce recurrent safety incidents and eliminate regulatory blind spots caused by default trust assumptions. To tackle the difficulty in predicting unconventional hazards and the lack of targeted control rules, the mechanisms of proactive prevention, continuous verification and dynamic adjustment inherent to Zero Trust support the construction of an adaptive risk management system (Figure 1). In cutting-edge experimental scenarios without mature regulatory standards, dynamic risk assessment and real-time intervention mechanisms can mitigate unknown risks proactively.

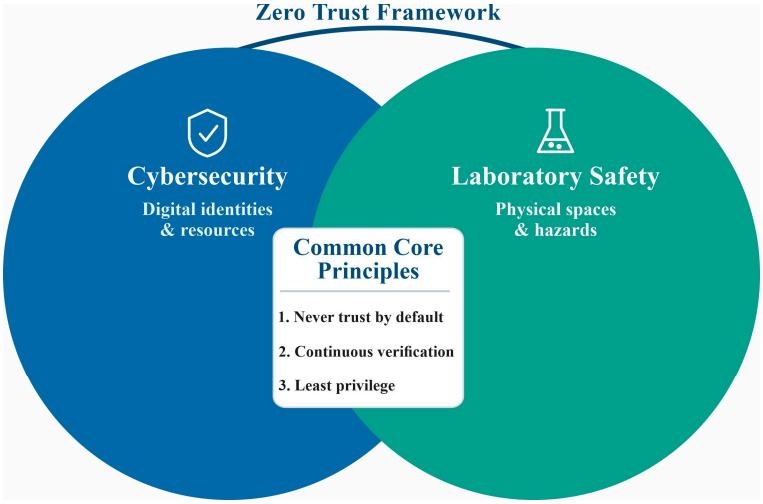


Figure 1. Cross-domain application of Zero Trust in cybersecurity and its adaptation to laboratory safety management.

Accordingly, the intertwined complexity and governance difficulties brought by both conventional and unconventional laboratory safety hazards render the conventional static rule-based model incompatible with the evolving landscape. Cross-domain introduction of Zero Trust represents a viable pathway to upgrade laboratory safety management from compliance-driven governance to intrinsic safety-oriented governance.

2.3. Core Characteristics of Laboratory-Oriented Zero Trust

Based on the Zero Trust concept originating from cybersecurity, this paper proposes a laboratory Zero Trust security philosophy adapted for university laboratory safety management. Abandoning the default assumption of inherent safety, this philosophy treats all core safety elements—laboratory personnel, mechanical equipment, experimental supplies, experimental methods and environmental conditions—as potential risk sources. Guided by the core principle of "never trust, always verify," this philosophy exhibits three core attributes:

First is universal distrust and dynamic verification. All safety elements in laboratories are deemed to carry potential safety risks, and no element is granted unconditional trust by default. Dynamic verification of pre-experiment qualifications and in-process operational behaviors is applied to all experimental activities. By collecting real-time safety status data of each core element, the regulatory loopholes generated by default trust in conventional management modes can be eliminated.

Second is least privilege and dynamic adjustment. Laboratory staff are only granted resource access permissions commensurate with their ongoing experimental activities;

such permissions will be automatically and dynamically adjusted according to changes in experimental project attributes, research cycles, risk levels and other factors, to mitigate safety risks induced by excessive access privileges.

Third is continuous monitoring and real-time early warning. Uninterrupted full-cycle monitoring is applied to personnel operations, equipment operating status, environmental parameters, chemical reagent consumption and other experimental conditions. Once monitoring indicators deviate from preset safety thresholds, the system will automatically trigger alarms, implement intervention measures and terminate experimental operations after feedback, realizing proactive risk prevention and control.

Compared with the conventional laboratory safety management model based on static rule restrictions, Zero Trust enables a fundamental shift in laboratory safety governance logic, namely shifting from passive post-accident disposal to proactive full-process monitoring and control, thus addressing the inherent limitations of conventional management including default trust, fixed access permissions and delayed feedback adjustment.

3. Core Implementation Priorities of Laboratory-Oriented Zero Trust

3.1. Construction of Dynamic Trust Mechanisms

Trust is not only the operational foundation of cybersecurity, but also a cornerstone of laboratory safety risk management, which directly shapes the fulfillment of safety responsibilities and the effectiveness of risk control outcomes. Zero Trust does not mean absolute distrust; it represents conditional dynamic trust, the essence of which is to establish a new trust mechanism supported by continuous verification. Zero Trust requires that no link of experimental activities be granted inherent safety credentials, and that all stages must undergo full-process dynamic compliance verification.

In the laboratory context, "trust" refers to the default assumption that people, equipment, materials and procedures will behave as expected and can therefore be exempted from repeated checking. Conventional management conserves effort by converting this one-time trust into a standing pass—once a person is trained or a facility certified, further verification is relaxed. The arrival of AI research weakens this assumption in two ways: on one hand, AI-assisted and agentic systems accelerate experiments and change operating states faster than human paperwork can re-verify; on the other, AI actions themselves are a new source of uncertainty, because an autonomous collaborator may pursue an experiment in ways no supervisor anticipated. Standing trust thus becomes a liability, and conditional, continuously re-verified trust becomes necessary.

3.2. Dynamic Control Based on the Least Privilege Principle

Permission management constitutes a critical component in laboratory safety prevention and control. Conventional management models are plagued by fixed access rights, insufficient review processes and delayed permission revocation. Permission management under the Zero Trust architecture is governed by the principles of least privilege and dynamic adjustment.

3.3. Development of Continuous Verification and Early Warning Systems

Conventional static rule-based laboratory safety management is prone to limitations such as delayed risk detection and slow emergency response. Zero Trust establishes a collaborative mechanism of continuous monitoring and real-time early warning, leveraging a suite of technologies including Internet of Things (IoT) monitoring, behavior recognition and analysis, environmental sensing and other technologies to realize real-time full-process supervision of personnel operational compliance, equipment operation status, environmental conditions and experimental material usage during experiments. When

monitoring indicators exceed predefined safety thresholds, the system automatically activates warning procedures and terminates dangerous operations, effectively addressing two major governance challenges: recurrent conventional hazards and unpredictability of unconventional risks.

Human factors remain a permanent source of risk: operators make mistakes, cut corners under time pressure, or drift from approved procedures, and even a well-designed system cannot fully eliminate such weaknesses. Continuous monitoring and AI-based recognition can help here by detecting anomalous conduct, near-misses and threshold breaches that human spot checks miss, and by issuing early warnings before an incident develops. At the same time, AI itself is not risk-free: model errors, sensor noise, accidental misuse, and the opacity of automated decisions can produce false reassurance or inappropriate intervention. For this reason we argue that humans should remain in the decision loop, but that the loop should be tiered rather than uniform: routine, low-risk judgments (for example, whether a permitted operation stays within a narrow threshold band) can be delegated to automated AI-driven decisions, whereas high-risk, high-consequence decisions—suspending an experiment, overriding an alarm, authorizing access to a new hazard—must retain human judgment. This tiered design both protects efficiency and preserves accountable human control.

4. Challenges in Implementing the Laboratory-Oriented Zero Trust

4.1. Balancing Safety, Research Efficiency and Operating Costs

The adoption of Zero Trust presents a core trade-off: aligning safety objectives with research efficiency gains and cost investment. At the safety governance level, real-time dynamic identity authentication and operation verification mechanisms may incur short-term efficiency losses; however, long-term application of intelligent dynamic assessment will enhance overall audit accuracy and comprehensive management efficacy. In terms of capital expenditure, the funds required for building Zero Trust and its supporting new technical systems will impose considerable financial pressure on a subset of laboratories. Therefore, the construction of Zero Trust and its supporting technical architecture must be guided by a coordinated assessment of safety, efficiency, and investment, and be rolled out flexibly according to differentiated practical demands of various laboratories. A related concern is that repeated verification may add time to already intensive interdisciplinary work and erode research efficiency. Rather than applying equally heavy verification to every activity, this overhead should be allocated by risk: high-risk experiments and unusual access requests undergo strict, multi-factor verification, while routine, low-risk operations rely on lightweight, largely automated checks that impose little visible delay. Because most verification steps can be performed in the background by sensors and intelligent platforms, the additional burden on researchers is expected to be small relative to the gains from earlier warning and fewer interruptions after incidents. In this way the system protects safety without uniformly slowing the research process.

4.2. Technical Barriers to Accurate Quantification of Dynamic Risks

As the core execution mechanism of Zero Trust, dynamic verification derives its practical efficacy from accurate risk quantification. A systematic indicator system for dynamic risk quantification and trust scoring must be developed to convert qualitative risk descriptions into quantifiable calculation parameters to provide relatively objective judgment criteria for dynamic risk assessment. Meanwhile, targeted research and development of quantitative prediction models for unconventional safety hazards is also needed, alongside quantitative analysis of the compounding effect of conventional and unconventional risks. Such methodological and technical obstacles will directly undermine the implementation

efficacy of Zero Trust security architecture. A workable metric system is expected to draw on several complementary dimensions—identity and qualification of the operator, real-time behavior and operational state, equipment and environmental parameters, and historical safety records—and to convert these into a transparent risk score with calibrated thresholds for normal, warning and intervention states. Because unconventional hazards are by definition poorly understood, these thresholds cannot be fixed in advance; they will require ongoing calibration, and the system should retain indicators whose performance can be audited. Establishing such a metric system is a prerequisite rather than a finished artifact, and its specification is an important task for future empirical work.

4.3. Compatibility Issues with Existing Static Management Systems

The laboratory safety responsibility system in universities is grounded in static responsibility delineation, which is partially misaligned with the dynamic governance requirements of Zero Trust. In terms of responsibility identification, fixed responsibility attribution modes struggle to accommodate dynamically adjusted access permissions under the Zero Trust framework, which can blur accountability boundaries. For assessment mechanisms, the current periodic evaluation model is temporally misaligned with the real-time data monitoring and dynamic evaluation emphasized by Zero Trust, hindering the alignment of the new Zero Trust-guided safety governance system with existing evaluation mechanisms. Furthermore, regarding management workflows, existing closed-loop management mechanisms may fail to respond promptly to violations and abnormal risks identified by the Zero Trust system. An effective linkage mechanism aligned with Zero Trust's proactive prevention philosophy must be established, to achieve dynamic closed-loop laboratory safety management.

5. Implementation Strategies for Laboratory-Oriented Zero Trust

5.1. Coordinated Implementation Framework of Institutional Improvement, Technical Support and Concept Renewal

The systematic rollout of Zero Trust and its supporting architecture demands holistic top-level design. In terms of concept popularization and training, special training and publicity campaigns should be conducted for laboratory safety supervisors at all levels, laboratory directors and frontline experimental staff, to deepen their understanding and acceptance of the core connotation and principles of Zero Trust security. At the institutional construction level, Zero Trust principles and supporting architectures should be fully integrated into existing institutional systems. For technical deployment, the technical compatibility of various IoT monitoring equipment, intelligent control systems and data management platforms should be systematically evaluated to reduce construction costs and technical difficulties.

5.2. Construction of Dynamic Intelligent Monitoring and Decision-Making Platforms

Guided by the core principle of "never trust, always verify", an intelligent monitoring platform covering all core laboratory safety elements should be developed to enable real-time dynamic data acquisition. Integrated with artificial intelligence technologies, the platform performs dynamic risk quantification and trust scoring, and implements tiered early warning and response measures according to scoring results. Taking the approval process of scientific research experimental projects as an example, a new workflow can be devised by integrating Zero Trust into conventional static rule management: after project review materials are submitted, multi-factor compliance verification is performed to complete static regulatory verification; artificial intelligence technologies are then adopted for dynamic trust scoring; finally, the risk decision-making module generates the final

review outcome, and the project can only be launched after passing approval. Centered on Zero Trust logic, this workflow integrates regulatory compliance constraints and dynamic trust assessment, enabling multi-dimensional, full-coverage review and intelligent, efficient decision-making.

5.3. Collaborative Integration with Existing Safety Management Systems

Following the principle of complementary advantages, integrated application of Zero Trust and the existing static rule-based management system shall be promoted. In terms of responsibility systems, the dynamic governance requirements of Zero Trust shall be effectively embedded into the current safety responsibility system, defining full-cycle dynamic management responsibilities for principal investigators of experimental projects and establishing a mechanism where management responsibilities adjust synchronously with access permissions. For reward, punishment and evaluation mechanisms, real-time monitoring data generated by the Zero Trust system shall be combined with conventional periodic evaluation indicators to build a comprehensive evaluation model integrating dynamic real-time data and static fixed indicators. In institutional coordination, a systematic review of incompatibilities between existing safety management systems and Zero Trust should be conducted, highlighting the core attributes of Zero Trust including universal distrust, dynamic verification, least privilege, dynamic adjustment, continuous monitoring and real-time early warning.

6. Future Perspectives

In the AI era, university laboratory safety management faces a complex situation with intertwined conventional and unconventional hazards, and the conventional static rule-based management model has demonstrated notable limitations. This study introduces Zero Trust, originating from cybersecurity, into university laboratory safety management via cross-domain adaptation, analyzes its core ideology and the essential principle of "never trust, always verify", and preliminarily discusses the core implementation priorities, practical obstacles and targeted promotion strategies of laboratory-oriented Zero Trust. Taken together, the proposed framework consists of three adapted attributes (universal distrust with dynamic verification, least privilege with dynamic adjustment, and continuous monitoring with real-time early warning), three implementation priorities, and a staged, tiered rollout that keeps human judgment for high-risk decisions; it is offered as a conceptual framework rather than a validated model. Looking forward, Zero Trust is poised to become an important guiding philosophy for university laboratory safety management in the AI era. It is anticipated that this research direction will attract more attention from laboratory safety researchers and frontline management practitioners, fostering in-depth theoretical research and practical application in this field. As the concept is gradually piloted and operational data accumulate, subsequent work will move from conceptual argument to empirical evaluation and a more rigorous treatment of the risk-management process.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

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

References

1. Tang, J.; Xia, L.; Li, Z.; Huang, C. AI-researcher: Autonomous Scientific Innovation. *Adv. Neural Inf. Process. Syst.* **2026**, *38*, 9481–9520.
2. Burrola-Núñez, H.; Álvarez-Chávez, C.R.; Sánchez-Duarte, N.E.; Burgos-Hernández, M.; Rojas-Moraga, Á.O. Advancing Laboratory Safety: The IHSCE Tool for Integrated Health, Safety, Civil Protection, and Environmental Management in University Laboratories. *ACS Chem. Health Saf.* **2026**, *33*, 147–157. [CrossRef]
3. 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]
4. Kindervag, J.; Balaouras, S. No More Chewy Centers: Introducing the Zero Trust Model of Information Security. Available online: https://media.paloaltonetworks.com/documents/Forrester-No-More-Chewy-Centers.pdf (accessed on 4 June 2026).
5. Ward, R.; Beyer, B. Beyond Corp: A New Approach to Enterprise Security. *Mag. Usenix Sage* **2014**, *39*, 6–11.
6. Stafford, V.A. Zero Trust Architecture. Available online: https://nvlpubs.nist.gov/nistpubs/specialpublications/NIST.SP.800-207.pdf (accessed on 4 June 2026).
7. Jiang, K.; Lin, Z.; Gao, L. Exploring AI-driven Transformation in Management Paradigms for Recurrent Safety Hazards in University Laboratories. *Laboratories* **2025**, *2*, 9. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.