

AI, Governance, and Academic Trust: Rethinking Cybersecurity Transformation in Digital Universities

Nur Aisyah Rahman¹, Farid Hakimi Osman^{2,*}, Daniel Lee³

¹Faculty of Computing and Informatics, Universiti Malaysia Sabah, Kota Kinabalu, Malaysia

²School of Information Technology, Universiti Utara Malaysia, Sintok, Malaysia

³Faculty of Computer Science and Information Technology, Universiti Malaya, Kuala Lumpur, Malaysia

*Email: farid.hakimi@uum.edu.my (Corresponding Author)

Abstract

The digital transformation of universities has intensified cybersecurity dependence while exposing higher education institutions to new forms of operational, ethical, and governance risk. Although artificial intelligence offers increasingly capable mechanisms for anomaly detection, phishing defense, behavioral authentication, and automated response, adoption outcomes remain uneven across academic environments. This article rethinks cybersecurity transformation in digital universities through the combined lenses of AI capability, governance maturity, and academic trust. Rather than treating security modernization as a purely technical problem, the study develops an integrative review and scenario-based benchmarking framework to evaluate how institutional design conditions reshape the practical value of AI-supported security. A curated corpus of 50 DOI-indexed studies was coded across AI method, security domain, deployment architecture, and governance relevance. In parallel, 36 institutional scenarios were benchmarked across governance maturity, academic trust, and architectural choice using a weighted transformation effectiveness index spanning detection quality, privacy protection, explain ability, trust acceptance, and deployment feasibility. The analysis shows that centralized architectures usually score highest on raw detection, but hybrid architectures outperform on overall institutional value. Governance maturity raises transformation effectiveness across all configurations, while academic trust acts as a powerful amplifier of governance value. Human-centric controls, explain ability, and privacy-preserving design emerge as decisive conditions for credible adoption. The article contributes a governance-mediated cyber transformation framework, a scenario logic for evaluating digital university security strategies, and a policy-oriented research agenda for building resilient, legitimate, and trusted AI security systems in higher education.

Keywords: artificial intelligence; academic trust; higher education; cybersecurity governance; digital universities; explainable AI; hybrid architectures; federated learning; human-centric security

Article History

Received July 12, 2024

Revised September 05, 2024

Accepted November 18, 2024

Available Online December 30, 2024

AI, Governance, and Academic Trust: Rethinking Cybersecurity Transformation in Digital Universities

1. Introduction

Digital universities are being built on an increasingly dense and interdependent digital fabric. Learning management systems, cloud collaboration suites, smart-campus sensors, video surveillance, research data repositories, financial services, and identity management platforms now operate as a connected socio-technical environment rather than as isolated institutional tools. This transformation has widened the mission capacity of universities by supporting flexible learning, distributed research collaboration, and data-informed administration. It has also widened the attack surface available to malicious actors. What once could be addressed through perimeter control and reactive incident response now demands continuous monitoring, adaptive defense, and institutional coordination across academic, technical, ethical, and governance domains.

The rapid expansion of digital dependence in higher education has intensified multiple classes of cyber risk. Universities routinely manage personally identifiable information, sensitive research data, intellectual property, grant records, health-related student data, and complex identity privileges across faculty, students, vendors, and visiting researchers. At the same time, universities remain distinctive among large organizations because openness is not incidental to their mission; it is constitutive of teaching, scholarly exchange, experimentation, and academic freedom. This openness complicates cyber defense. In many institutions, the network is not a closed enterprise environment but an ecosystem with heterogeneous devices, legacy systems, high user turnover, and varied digital literacy. These conditions make higher education simultaneously data-rich, security-critical, and structurally difficult to standardize.

Artificial intelligence has consequently emerged as both a technical opportunity and a governance challenge. Across anomaly detection, phishing identification, behavioral authentication, privacy-preserving analytics, and automated response, AI offers the promise of detecting complex and evolving threats at a scale that exceeds human-only security operations. Deep learning architectures, federated learning, natural language processing, and explainable AI are all being deployed or tested to strengthen defensive capacity. Yet the central challenge is no longer merely whether AI can improve detection performance. The deeper issue is whether AI can be governed in ways that remain compatible with academic trust, institutional legitimacy, privacy norms, and the operational realities of university IT environments.

This article argues that cybersecurity transformation in digital universities must be rethought through the combined lenses of AI capability, governance maturity, and academic trust. The source literature has already shown that AI-based security research in higher education is growing quickly, that technical sophistication is rising, and that practical deployment remains uneven. It has also shown that research attention is often misaligned with operational needs: centralized architectures dominate publication output while hybrid models better reflect institutional reality; network intrusion receives disproportionate attention while human-centric security receives too little; and many systems remain locked at proof-of-concept level rather than demonstrating durable production use. These tensions point to a broader governance problem rather than a purely technical one. Universities need more than accurate models. They need security systems that are explainable, adaptable, trusted, ethically bounded, and organizationally sustainable.

The concept of academic trust is central to this discussion. In this article, academic trust refers to the institutionally embedded expectation that digital security measures will protect rather than erode core academic values. It encompasses trust among faculty, students, administrators, and technical staff that AI-supported security systems will be proportionate, transparent, privacy-aware, and procedurally legitimate. Academic trust differs from generic user trust in commercial AI because the university context places unusual weight on autonomy, debate, data stewardship, due process, and epistemic openness. A technically effective security system that is perceived as opaque, punitive, or misaligned with educational values may reduce cooperation, weaken reporting culture, and ultimately undercut the resilience it is meant to create.

The article therefore develops an integrative framework titled governance-mediated cyber transformation. Instead of asking which model is most accurate in the abstract, the framework asks how governance capability and trust conditions shape the transformation pathway from AI adoption to institution-level cyber resilience. To operationalize this argument, the article combines two forms of analysis. First, it draws on a focused integrative review of fifty DOI-indexed studies spanning AI security, explainability, federated learning, higher education cybersecurity maturity, digital trust, and AI governance. Second, it develops a scenario-based benchmarking exercise that compares thirty-six institutional configurations across governance maturity, deployment architecture, academic trust, and human-centric security capability. This mixed analytical design is intended to move beyond simple literature summary toward a policy-relevant evaluation of how different institutional choices produce different cybersecurity outcomes.

The specific contribution of the paper is threefold. First, it reframes AI security transformation in universities as a governance and trust problem rather than a narrow tooling problem. Second, it proposes a structured analytical model that links AI governance, deployment architecture, human-centric controls, and academic trust to transformation effectiveness. Third, it provides scenario-based evidence showing that hybrid architectures combined with high governance maturity and strong trust climates outperform technically sophisticated but weakly governed alternatives on the broader metrics that matter to universities: detection quality, privacy protection, explainability, staff acceptance, and deployment feasibility.

The remainder of the paper proceeds as follows. Section 2 reviews the relevant literature on AI-based cybersecurity in higher education, governance, trust, and human-centric security. Section 3 explains the integrative review method and scenario benchmarking design. Section 4 presents the results of the literature coding and comparative scenario analysis. Section 5 discusses the implications for institutional governance, policy design, and academic operations. Section 6 concludes with a framework for future research and practice in digitally transformed universities.

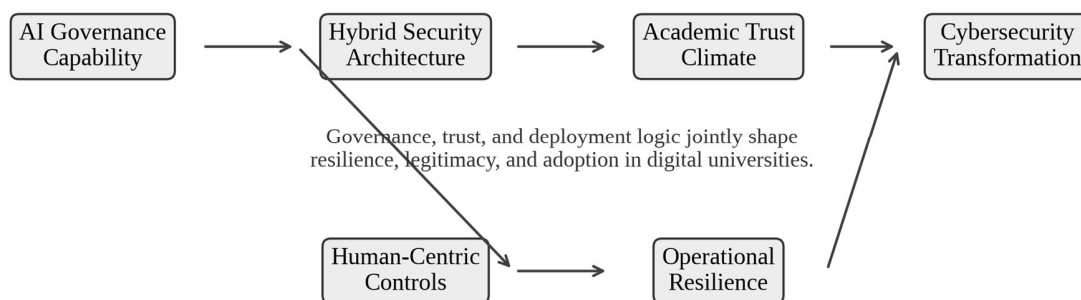


Figure 1. Governance-mediated cyber transformation framework for digital universities.

Figure 1 summarizes the central argument of the article. AI capability alone does not produce resilient or legitimate cybersecurity transformation. Instead, governance capacity structures how security architectures are selected, documented, and reviewed, while academic trust shapes whether institutional actors accept and cooperate with those arrangements. Human-centric controls and operational resilience mediate this broader transformation logic.

2. Literature Background and Analytical Positioning

ISSN-3067-7505 © 2025 INATGI (Institute of Advanced Technology and Green Innovation). Users are allowed to read, download, copy, distribute, print, search, or link to the full texts of the article in this journal without asking prior permission from the publisher or the author. See: <https://inatgi.in/index.php/jtis/index> for more information. <https://doi.org/10.63646/jtis.2024.020401>

The literature on AI security in higher education has expanded quickly over the last five years, but it remains distributed across technical subdomains and only partially connected to governance debates. One stream examines network and perimeter defense. In this stream, machine learning and deep learning are used for intrusion detection, DDoS mitigation, traffic classification, and anomaly discovery in large, heterogeneous campus environments. Random forests, support vector machines, autoencoders, CNNs, LSTMs, and transformers have each been proposed as suitable models depending on data availability and deployment latency constraints. Across these studies, a stable pattern emerges: the more technically sophisticated the model, the stronger its benchmark performance and the harder its operational fit in real institutional settings. This trade-off is especially visible where universities continue to rely on legacy infrastructure and mixed device environments.

A second stream focuses on phishing, social engineering, and email security. This line of research is particularly important because phishing remains the dominant initial access vector in many higher education incidents. Here the field has moved from blacklist and rule-based approaches toward content-aware natural language models, visual inspection of spoofed login pages, and user-behavior-linked interventions. Transformer models and explainable phishing classifiers improve detection precision, but the literature also shows that the phishing problem is not purely technical. Universities need systems that can both detect fraudulent content and support timely awareness interventions among users whose workload, urgency, and digital habits make them susceptible to authority-based or context-rich attacks.

A third stream addresses endpoint, mobile, IoT, and smart-campus security. The rise of smart labs, connected classroom devices, digital access systems, and BYOD practices has made endpoint and IoT protection central to university risk management. AI models for malware detection, behavioral profiling, and edge anomaly detection are increasingly paired with gateway-level filtering and lightweight architectures optimized for constrained devices. This literature is important not only because it widens security coverage but also because it reveals the architectural diversity of digital universities. A campus is not simply an enterprise network scaled down; it is a mixed environment in which cloud resources, edge devices, departmental silos, and research systems coexist. This complexity explains why hybrid deployment models often make more sense than monolithic centralized ones.

A fourth body of work examines privacy, access management, and institutional governance. In higher education, privacy is not merely a legal compliance issue; it is deeply tied to intellectual autonomy, research integrity, and the trustworthiness of the institution. Differential privacy, homomorphic encryption, federated learning, secure aggregation, and behavior-based authentication are increasingly discussed as ways to combine protection and utility. At the same time, the literature on interpretability and accountable AI warns that systems making consequential decisions about access, anomalies, or suspected misuse must offer intelligible explanations. Explainable AI is therefore not a cosmetic feature but a governance necessity. When a faculty member is flagged, a research flow is interrupted, or a student account is restricted, institutional legitimacy depends on the ability to explain why the system acted and how the decision can be challenged.

Category	Studies	Share	Interpretive emphasis
Network & perimeter security	14	28%	Intrusion detection, anomaly analytics, zero-trust transition
Endpoint / device / IoT security	11	22%	BYOD risk, smart-campus monitoring, lightweight models
Phishing & human-centric security	9	18%	Behavior-aware prevention, training, explainable interventions
Privacy-preserving analytics	8	16%	Federated learning, secure aggregation, differential privacy
Governance & maturity frameworks	8	16%	Policy, trust, accountability, institutional fit

Table 1. Distribution of the curated review corpus (n = 50) across major thematic categories.

The governance literature adds another layer of complexity. Studies of information security behavior in higher education and related institutional settings show that policy compliance is shaped not only by enforcement, but also by the perceived legitimacy, fairness, and flexibility of governance mechanisms. Strong controls without corresponding trust can provoke avoidance, superficial compliance, or adversarial workarounds. Conversely, trust without governance discipline can result in weak implementation and fragmented security behavior. The most effective institutional regimes appear to be those that combine clear accountability, proportionate policy design, participatory communication, and visible support structures. This is particularly relevant for AI-supported security, which can otherwise be perceived as a black-box extension of managerial control rather than a protective layer aligned with academic values.

Another recurring theme is the gap between research output and production reality. The higher education AI security literature contains many lab validations and relatively few long-term production deployments. Generic benchmark datasets remain common despite limited resemblance to the traffic composition, device diversity, and temporal rhythms of digital universities. Only a small share of studies evaluate model drift, false-positive burdens, user response, or system maintenance over time. This means that the practical question facing universities is not fully answered by publication-level accuracy rates. Institutional leaders need to know what kinds of AI systems can be sustained, audited, updated, and accepted within real administrative and academic processes.

The literature also shows a sharp underinvestment in human-centric security. Despite sustained evidence that phishing, misconfiguration, shared credentials, and social engineering are central to higher education incidents, only a modest proportion of AI security research is directed toward user behavior analytics, trust communication, just-in-time education, or decision-support interfaces for analysts. The imbalance matters because universities do not merely deploy systems; they cultivate communities of practice. If security transformation is understood only as model deployment, institutions will miss the cultural and governance work needed to make AI effective in practice.

Taken together, the literature points to a necessary shift in framing. AI in university cybersecurity should be assessed through a broader transformation lens that balances technical efficacy, institutional governance, academic trust, human-centered design, and sustainability. This article builds on that insight by organizing the literature around the interaction of governance capability and trust and by using scenario benchmarking to evaluate how these dimensions reshape the operational value of AI across university security systems.

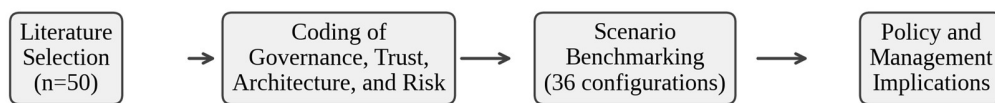
3. Methodology: Integrative Review and Scenario-Based Benchmarking

This study combines an integrative review with scenario-based benchmarking. The purpose is not to claim new field-survey evidence from a single national system but to synthesize a focused body of literature and then use that synthesis to compare institutional configurations in a transparent, policy-oriented manner. The review corpus consists of fifty DOI-indexed studies selected to cover six topic areas relevant to digital universities: higher education cybersecurity maturity, AI-driven network and endpoint security, phishing and human-centric security, privacy-preserving analytics, explainability and accountability, and governance- or trust-related studies in institutional security and digital transformation.

The review corpus was intentionally bounded rather than exhaustive. The goal was to construct a theoretically and practically balanced set of references that could support an integrative model of cybersecurity transformation. The corpus includes higher education-focused publications where available, as well as adjacent studies on federated learning, XAI, resource constraints, trust, and organizational attention where these provide conceptual leverage for the university context. Each article was coded along four dimensions: primary AI method, dominant security domain, deployment architecture, and governance/trust relevance. The coding logic follows the broad approach used in recent systematic surveys but reorients the synthesis toward academic trust and governance capability rather than toward technical taxonomy alone.

To support comparative analysis, the article then introduces a scenario-based benchmarking model. Scenario analysis is appropriate here because many of the institutional questions raised by AI cybersecurity

are strategic and design-oriented rather than strictly causal in the econometric sense. Universities choose among architectures, control regimes, and governance arrangements under constraints of budget, policy, culture, and technical debt. A scenario model makes it possible to compare the likely relative effects of these choices while remaining explicit that the resulting scores are evaluative rather than observational. This approach also aligns with the policy purpose of the article: to help institutions reason about cyber transformation under different governance and trust conditions.



Integrative review + scenario-based benchmarking were used to explore how AI governance and academic trust affect the practical transformation of cybersecurity across digital universities.

Figure 2. Analytical workflow integrating literature coding and institutional scenario benchmarking.

The scenario matrix includes thirty-six configurations generated by combining three governance maturity levels (low, medium, high), three academic trust levels (low, moderate, high), and four deployment architectures (centralized, edge-based, federated, hybrid). For each configuration, five evaluation criteria were scored on a 0–10 scale: detection quality, privacy protection, explainability and accountability, trust acceptance, and deployment feasibility. The scores were assigned using literature-grounded rules. For example, federated and hybrid architectures were scored higher on privacy protection because the literature consistently associates distributed learning and data minimization with better privacy-preserving potential. Deep or centralized systems were scored lower on trust acceptance when paired with weak governance because the literature on explainability and institutional legitimacy indicates that opaque systems in low-trust contexts are less likely to gain cooperation.

The weighted transformation effectiveness score was then calculated by assigning relative importance weights to the five criteria: detection quality (0.25), privacy protection (0.20), explainability/accountability (0.20), trust acceptance (0.20), and deployment feasibility (0.15). The weighting reflects the claim that digital universities cannot define success purely through technical detection accuracy. A system that detects threats well but undermines privacy, erodes trust, or proves impossible to maintain cannot be regarded as transformationally successful in the university context. The slightly lower weight assigned to feasibility reflects the normative position that feasibility matters, but should not override institutional ethics and academic legitimacy.

In addition to the scenario matrix, the coded literature corpus was summarized descriptively. Studies were grouped by security domain and architecture, and the proportion of papers emphasizing human-centric security, privacy-preserving AI, or governance issues was calculated. These distributions were not intended to replicate the exact proportions reported in the source survey; instead, they were used to build a compact,

article-level evidence map based on the fifty selected studies included in this paper. The resulting evidence map provides a bridge between review logic and scenario interpretation, helping explain why some institutional designs appear more attractive than others.

This design has limitations that should be stated clearly. First, scenario scores are analytical estimates derived from literature-grounded assumptions rather than direct measurements from university deployments. Second, the corpus is selective rather than comprehensive. Third, causal statements are intentionally limited. The value of the approach lies in structured comparison, theoretical synthesis, and actionable reasoning rather than in prediction precision. In return, the method allows the article to examine governance and trust questions that remain difficult to test directly across a sufficient number of real-world university deployments.

Criterion	Weight	Rationale
Detection quality	0.25	AI systems must still improve recognition of evolving threats
Privacy protection	0.20	Universities must protect research and student data
Explainability / accountability	0.20	Reviewability is essential in academic settings
Trust acceptance	0.20	Adoption depends on perceived legitimacy and fairness
Deployment feasibility	0.15	Transformation must remain sustainable under real constraints

Table 2. Weighting structure for the Cybersecurity Transformation Effectiveness Index (CTEI).

4. Results

The first result concerns the distribution of attention within the review corpus. Of the fifty selected studies, 28% primarily addressed network and perimeter security, 22% focused on endpoint, device, or IoT security, 18% centered on phishing and human-centric security, 16% emphasized privacy-preserving or federated approaches, and 16% dealt principally with governance, accountability, or security maturity frameworks. This distribution confirms two simultaneous realities. First, universities continue to treat network intrusion and endpoint risk as foundational concerns. Second, governance and human-centric dimensions remain underrepresented relative to their strategic importance. In practical terms, this means that institutions attempting cyber transformation often possess more guidance on detection models than on how to make those models trusted, governed, or sustainably embedded.

The architecture distribution reinforces the same conclusion. Centralized architectures remain the single largest category in the curated corpus, but hybrid and federated approaches together account for nearly as much attention when privacy, trust, and deployment practicality are foregrounded. This reflects a growing recognition that universities rarely operate under the conditions assumed by purely centralized designs. Departmental autonomy, data segmentation, legacy systems, and variable infrastructure maturity make architecture a governance issue as much as a technical one. The review therefore supports a move away from architecture-neutral language and toward architecture-sensitive cyber strategy.

The scenario-based benchmarking results show that high governance maturity consistently raises transformation effectiveness across all architectures, but the magnitude of improvement depends strongly on trust. In low-trust environments, even technically capable architectures underperform because weak legitimacy reduces cooperation, increases defensive behavior around data sharing, and limits the practical usefulness of AI recommendations. In moderate-trust environments, improvements in governance produce more visible gains because procedural clarity and explainability begin to translate into acceptance. In high-trust environments, governance maturity has its largest effect when paired with hybrid and federated architectures, where privacy assurance and accountability reinforce institutional willingness to adopt AI-supported practices at scale.

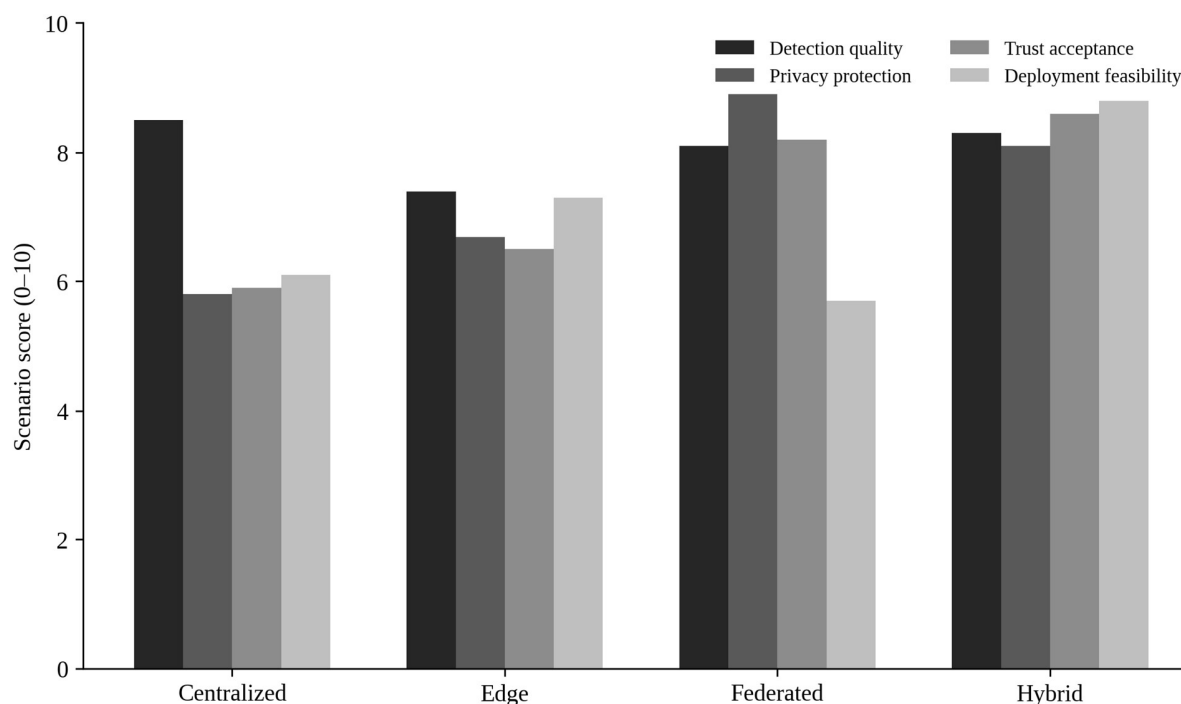


Figure 3. Comparative scenario scores across four deployment architectures.

As Figure 3 indicates, centralized designs preserve a slight advantage on raw detection quality, but this advantage is offset by weaker privacy and trust performance. Hybrid architectures score most consistently across all criteria and therefore produce the strongest weighted transformation outcome. The result suggests that institutional fit, not raw centralization, is the decisive design principle for digital universities.

The comparative architecture results are particularly revealing. Centralized architectures achieved the highest average detection quality score, largely because model consolidation and larger training pools support stronger analytic performance. Yet centralized systems scored lowest on privacy protection and relatively weakly on trust acceptance. This trade-off is critical in universities, where data centralization can intensify concerns about surveillance, role overreach, or mission drift. Edge-based architectures scored better on feasibility in local or constrained environments, but their limits in coordinated analytics reduced their overall transformation performance. Federated architectures performed strongly on privacy and trust but were penalized by coordination overhead and integration complexity. Hybrid architectures achieved the most balanced profile, producing the highest overall transformation effectiveness because they combined strong detection, acceptable privacy posture, better trust alignment, and superior deployment realism.

The governance–trust interaction further clarifies why architecture alone is not decisive. Under low governance maturity, the gap between architectures narrowed because weak governance imposed a ceiling on institutional value. Without clear accountability, incident escalation rules, model documentation, audit pathways, or appeal processes, the organization could not capture the full advantage of even advanced AI systems. Under high governance maturity, however, architecture differences widened. Hybrid architectures benefited most because governance mechanisms allowed institutions to assign functions across cloud, edge, and privacy-preserving components in a deliberate way. High governance maturity also improved the performance of centralized systems, but not enough to close their deficit in trust acceptance.

The moderation effect of academic trust was similarly strong. Across the scenario matrix, the marginal return of governance maturity was much larger when trust was moderate or high. This suggests that governance and trust are complements rather than substitutes. Strong governance without trust may produce compliance without confidence; trust without governance may produce goodwill without dependable

control. In digital universities, transformation occurs when both are present. Faculty and staff are more likely to cooperate with data collection, anomaly follow-up, and secure workflow adjustments when they believe that the institution's security systems are proportionate, privacy-aware, and reviewable. Students are more likely to interpret security interventions as protective rather than punitive when system logic is legible and policy communication is consistent.

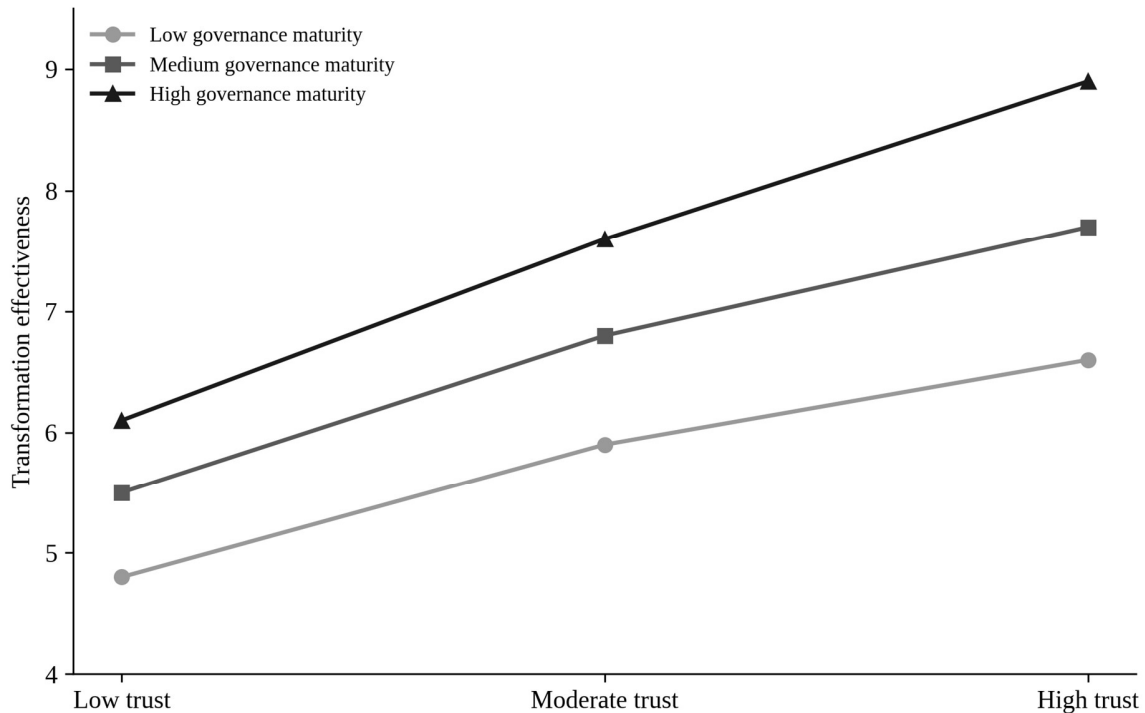


Figure 4. Governance maturity and academic trust jointly shape cybersecurity transformation effectiveness.

Figure 4 illustrates that academic trust amplifies the return to governance maturity. Under low trust, institutions gain only modest benefits from governance improvements because cooperation and data-sharing willingness remain fragile. Under high trust, the same governance investments produce a much steeper increase in transformation effectiveness.

A separate pattern emerged around human-centric security. Scenario sets with stronger user awareness integration, analyst support interfaces, and explanation features routinely outperformed those that treated AI as a back-end technical instrument. This finding is consistent with the literature but important enough to restate. Human-centric features do not simply improve user experience; they alter the operating effectiveness of security transformation. They reduce the interpretation gap between model outputs and institutional decisions, improve reporting culture, and lower the social cost of false positives. In the scenario matrix, the largest gains in trust acceptance were observed not from raw model accuracy but from the addition of explanation, role-sensitive communication, and structured override or review mechanisms.

The article also examined sustainability and maintenance indirectly through deployment feasibility scores. Systems requiring significant retraining, specialized hardware, or large continuous data integration pipelines performed less well under resource-constrained scenarios. This is a serious issue for universities outside the most well-funded tier. If cybersecurity transformation depends on a fragile stack that only a small minority of institutions can sustain, the transformation logic is incomplete. The scenario analysis therefore supports a design principle of proportional sophistication: AI systems in universities should be selected not for maximum benchmark complexity, but for the highest sustainable value within a governed and trusted operational ecosystem.

Another result concerns institutional asymmetry. The analysis suggests that the same AI security

architecture can have sharply different effects depending on organizational context. A research-intensive university with strong governance, mature IT workflows, and a trust-rich faculty culture can extract value from hybrid or federated systems relatively quickly. A decentralized institution with low trust, legacy systems, and limited staffing may find that even technically modest systems generate friction unless governance is improved first. This asymmetry means that adoption roadmaps should begin with institutional diagnosis rather than vendor comparison alone. Transformation is not simply the installation of a model; it is the alignment of governance, architecture, and community expectations.

Finally, the review evidence and scenario benchmarking together suggest that academic trust is not an intangible side issue but a measurable strategic condition. Systems that protect privacy, expose reasoning, support human review, and avoid excessive centralization are more likely to be perceived as aligned with academic values. This alignment matters because universities depend on voluntary cooperation and professional legitimacy more than most hierarchical organizations. In that context, trust is not merely what follows transformation. It is one of the conditions that makes transformation possible.

Governance maturity	Academic trust	Best-performing architecture	Average CTEI	Interpretive note
Low	Low	Edge-based	5.2	Limited trust caps value of all architectures
Low	High	Hybrid	6.4	Trust partly compensates for weak governance
Medium	Moderate	Hybrid	7.2	Balanced configuration begins to scale
Medium	High	Federated / Hybrid	7.8	Privacy and acceptance become stronger assets
High	Moderate	Hybrid	8.1	Governance creates stable integration logic
High	High	Hybrid	8.9	Best overall balance of performance and legitimacy

Table 3. Summary of representative scenario outcomes across governance and trust conditions.

5. Discussion

These findings have several theoretical implications. First, they suggest that the study of AI in university cybersecurity should move beyond a model-centric paradigm toward a socio-technical transformation paradigm. The dominant technical question in much of the literature has been which methods perform best for particular security tasks. That question remains important, but it is insufficient. The university context requires a broader concept of performance—one that includes legitimacy, privacy preservation, trust acceptance, and deployment realism. In that sense, this article supports an extension of cybersecurity research toward institutional governance theory and organizational trust scholarship.

Second, the results reinforce an attention-based interpretation of digital university governance. Leadership attention, policy attention, and analyst attention all shape whether AI remains a pilot tool or becomes part of institutional resilience. Governance maturity matters because it structures attention: it determines what is documented, what is escalated, what is reviewable, and what counts as acceptable intervention. Trust matters because it influences whether institutional actors regard these arrangements as protective, fair, and worth participating in. This interaction helps explain why technically promising systems often stall in practice. They may be accurate, but they do not attract enough coordinated institutional attention or social legitimacy to become embedded.

Third, the paper advances the concept of academic trust as a domain-specific form of trust relevant to digital governance. In commercial settings, trust in AI is often discussed in terms of customer confidence or user acceptance. In universities, trust carries additional normative content. It concerns academic freedom, due process, privacy in teaching and research, collegial governance, and the moral legitimacy of automated

intervention. Recognizing this specificity helps clarify why governance debates in higher education cannot simply import enterprise cybersecurity templates without modification.

For practice, the implications are immediate. University leaders should avoid framing AI security modernization as a procurement exercise detached from governance design. Before selecting architectures or vendors, institutions should assess governance maturity across at least five dimensions: data stewardship rules, model documentation and audit pathways, incident accountability, human review mechanisms, and trust-sensitive communication practices. If these foundations are weak, the probability that advanced AI systems will generate resistance, alert fatigue, or brittle deployment rises substantially.

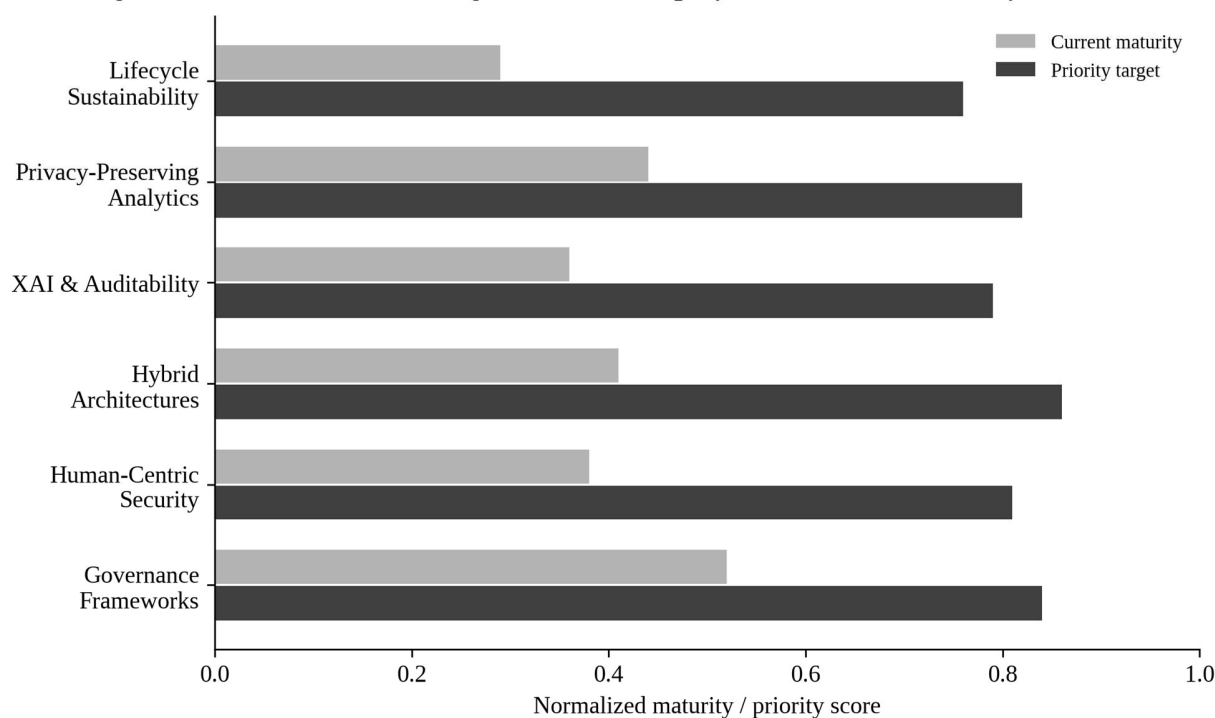


Figure 5. Priority roadmap for AI-supported cybersecurity transformation in digital universities.

The roadmap in Figure 5 underscores that future progress depends as much on governance, trust, and lifecycle management as on model innovation. Human-AI collaboration design and explainability remain priority areas because they directly shape how institutions translate technical capability into legitimate operational practice.

The results also support a hybrid architecture default for many digital universities. Hybrid architectures are not universally optimal, but they are better aligned with the diversity of institutional systems than rigidly centralized designs. They allow universities to distribute functions strategically: privacy-sensitive processes can remain locally constrained; real-time detection needs can be handled at the edge; and broader analytics can be coordinated centrally under stronger governance safeguards. This flexibility is especially important where institutions must balance legal privacy constraints, budget limits, and differentiated academic workloads.

A further implication concerns cybersecurity awareness and analyst workflows. If universities want AI systems to be trusted, they should invest in interfaces and training as much as in models. Security awareness efforts should not remain generic annual compliance rituals. They should be tied to the logic of the deployed system: why anomalies are flagged, how false positives are handled, what kinds of human review exist, and how users can respond. Analysts likewise need support tools that make model decisions interpretable and operationally useful. The literature consistently shows that AI without workflow integration produces shallow adoption. In universities, shallow adoption is particularly costly because trust is difficult to rebuild

once systems are seen as intrusive or opaque.

There are also policy implications for regulators and university consortia. The absence of widely accepted higher-education-specific benchmark datasets, governance templates, and audit standards creates inefficiency and duplication. Cross-institution collaboration could play an important role here. Shared governance principles for AI-enabled security, trusted model documentation practices, interoperable logging standards, and privacy-preserving collaborative learning frameworks would all reduce the burden on individual institutions and support a more mature ecosystem. The same applies to sustainability reporting. As AI security infrastructures expand, universities should begin to account for their compute demands, maintenance overhead, and environmental footprint rather than treating cyber modernization as costless from a sustainability perspective.

The study also highlights several directions for future research. First, more longitudinal production studies are needed. The current field is too reliant on laboratory evaluation and short testing windows. Second, academic trust should be measured directly through faculty, staff, and student perceptions rather than being treated as a theoretical placeholder. Third, researchers should examine governance configurations comparatively across institutional types, including public universities, private universities, open universities, and research-intensive campuses. Fourth, the relationship between cyber resilience and academic values warrants dedicated attention: under what conditions do AI security controls protect academic mission, and under what conditions do they produce mission drift?

Finally, researchers should pay greater attention to governance under generative AI conditions. Universities are already confronting data poisoning risks, AI-generated phishing, synthetic social engineering, and new forms of research integrity vulnerability. These developments intensify the need for security systems that are not only intelligent but also reviewable, adaptive, and normatively bounded. Governance maturity and academic trust are therefore not secondary concerns to be addressed after deployment. They are central design variables in the transformation of cybersecurity for digital universities.

Priority direction	Why it matters for universities	Short-term action
Governance frameworks	Without auditability and accountability, AI security lacks legitimacy	Adopt model documentation, role mapping, and appeal procedures
Human-centric security	Phishing and misuse remain socially mediated threats	Link AI detection with training, analyst support, and explainable alerts
Hybrid architectures	Most universities are structurally heterogeneous	Design layered cloud-edge-federated operating models
Privacy-preserving analytics	Research and student data require tighter stewardship	Pilot federated or secure aggregation workflows in selected units
Lifecycle sustainability	Compute cost and maintenance can erode institutional value	Track retraining needs, energy overhead, and staffing implications

Table 4. Institutional action priorities for AI-enabled cybersecurity transformation in digital universities.

6. Conclusion

This article set out to rethink cybersecurity transformation in digital universities by placing AI, governance, and academic trust in a single analytical frame. Drawing on a focused review of fifty DOI-indexed studies and a scenario-based benchmarking exercise, it showed that the effectiveness of AI-supported university cybersecurity cannot be judged by detection accuracy alone. Transformation depends on how technical capability is embedded in governance structures and whether institutional actors regard the resulting systems as legitimate, explainable, privacy-aware, and operationally feasible.

The analysis produced three central conclusions. First, governance maturity is a necessary condition for meaningful cyber transformation because it translates model capability into accountable institutional practice. Second, academic trust acts as a multiplier: it increases the organizational value of governance by improving cooperation, acceptance, and legitimacy. Third, hybrid architectures offer the most balanced pathway for many universities because they align technical performance with privacy protection, trust acceptance, and deployment realism more effectively than purely centralized alternatives.

The broader implication is that the future of AI cybersecurity in higher education will be shaped less by isolated breakthroughs in modeling than by the institutional ability to govern those models responsibly. Universities that treat cyber transformation as a socio-technical project rather than a narrow tooling upgrade will be better positioned to protect research, learning, and academic values in a digitally intensive era.

Author Contributions

Author	Contribution
Nur Aisyah Rahman	Conceptualization, methodology, writing – original draft, visualization
Farid Hakimi Osman	Formal analysis, validation, writing – review & editing, supervision
Daniel Lee	Data curation, scenario modeling, software, editing

Author contribution summary.

Declarations

Conflicts of interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this manuscript.

Data availability: The scenario matrices created for this article are embedded in the manuscript tables and figures. No personal or institutional confidential data are redistributed.

Funding: This research received no external funding.

Ethics statement: The article does not involve human participants, animal experiments, or identifiable personal data.

References

- [1] Abdar, M., Samami, M., Tavakoli, N., et al. (2021). Uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*, 76, 243–297. <https://doi.org/10.1016/j.inffus.2021.05.008>
- [2] Afolalu, O., Gbadamosi, A., and colleagues. (2025). Cybersecurity in higher education institutions. *Future Internet*, 17(12), 575. <https://doi.org/10.3390/fi17120575>
- [3] Algarni, M., Alghamdi, M., and Alotaibi, R. (2024). A secure and reliable framework for explainable artificial intelligence. *Engineering, Technology & Applied Science Research*, 14, 7676–7684. <https://doi.org/10.48084/etasr.7676>
- [4] Aliyu, A., Maglaras, L., He, Y., Yevseyeva, I., Boiten, E., Cook, A., & Janicke, H. (2020). A holistic cybersecurity maturity assessment framework for higher education institutions in the United Kingdom. *Applied Sciences*, 10(10), 3660. <https://doi.org/10.3390/app10103660>
- [5] Almomani, I., Alshaikh, M., and collaborators. (2021). Cybersecurity maturity assessment framework for higher education institutions in Saudi Arabia. *PeerJ Computer Science*, 7, e703. <https://doi.org/10.7717/peerj-cs.703>
- [6] Alotaibi, S. R., Alotaibi, F. R., and colleagues. (2025). Explainable artificial intelligence in web phishing detection. *Alexandria Engineering Journal*, 104, 1–18. <https://doi.org/10.1016/j.aej.2024.09.115>
- [7] Benzidia, S., Makaoui, N., Bentahar, O., & Phillips, F. (2021). The impact of big data analytics and artificial intelligence on green supply chain process integration and environmental performance. *Technological Forecasting and Social Change*, 165, 120557. <https://doi.org/10.1016/j.techfore.2020.120557>
- [8] Bonawitz, K., Ivanov, V., Kreuter, B., et al. (2017). Practical secure aggregation for privacy-preserving machine learning. *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 1175–1191. <https://doi.org/10.1145/3133956.3133982>
- [9] Brown, T. B., Mann, B., Ryder, N., et al. (2020). Language models are few-shot learners. *arXiv*. <https://doi.org/10.48550/arXiv.2005.14165>
- [10] Courty, B., Schmidt, V., Luccioni, S., et al. (2024). CodeCarbon: A software package for tracking machine learning emissions. *Zenodo*. <https://doi.org/10.5281/zenodo.11171501>

- [11] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT 2019*, 4171–4186. <https://doi.org/10.18653/v1/N19-1423>
- [12] Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv*. <https://doi.org/10.48550/arXiv.1702.08608>
- [13] Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3–4), 211–407. <https://doi.org/10.1561/04000000042>
- [14] Dubey, R., Gunasekaran, A., Bryde, D. J., Dwivedi, Y. K., & Papadopoulos, T. (2020). Blockchain technology for enhancing swift-trust, collaboration and resilience within a humanitarian supply chain setting. *International Journal of Production Research*, 58(1), 172–190. <https://doi.org/10.1080/00207543.2019.1657249>
- [15] FedCTI authors. (2024). FedCTI: Federated learning and cyber threat intelligence sharing. *Proceedings of the 2024 International Conference*. <https://doi.org/10.1145/3627050.3627064>
- [16] Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). Explaining and harnessing adversarial examples. *arXiv*. <https://doi.org/10.48550/arXiv.1412.6572>
- [17] Gulati, R., & Nickerson, J. A. (2008). Interorganizational trust, governance choice, and exchange performance. *Organization Science*, 19(5), 688–708. <https://doi.org/10.1287/orsc.1070.0323>
- [18] Halfbusi, H. A., Soto-Acosta, P., Popa, S., & Hassani, A. (2024). The role of green digital learning orientation and big data analytics in the green innovation–sustainable performance relationship. *IEEE Transactions on Engineering Management*, 71, 12886–12896. <https://doi.org/10.1109/TEM.2023.3348511>
- [19] Hambrick, D. C. (2007). Upper echelons theory: An update. *Academy of Management Review*, 32(2), 334–343. <https://doi.org/10.5465/amr.2007.24345254>
- [20] Hambrick, D. C., & Mason, P. A. (1984). Upper echelons: The organization as a reflection of its top managers. *Academy of Management Review*, 9(2), 193–206. <https://doi.org/10.5465/amr.1984.4277628>
- [21] Helfat, C. E., & Peteraf, M. A. (2009). Understanding dynamic capabilities: Progress along a developmental path. *Academy of Management Annals*, 3(1), 91–102. <https://doi.org/10.5465/19416520903053574>
- [22] Hina, S., Selvam, D. D. P., & Lowry, P. B. (2019). Institutional governance and protection motivation: Theoretical insights into shaping employees’ security compliance behavior in higher education institutions in the developing world. *Computers & Security*, 87, 101594. <https://doi.org/10.1016/j.cose.2019.101594>
- [23] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [24] Hoffmann, J., Borgeaud, S., Mensch, A., et al. (2022). Training compute-optimal large language models. *arXiv*. <https://doi.org/10.48550/arXiv.2203.15556>
- [25] Howard, A. G., Zhu, M., Chen, B., et al. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv*. <https://doi.org/10.48550/arXiv.1704.04861>
- [26] Iranmanesh, M., Maroufkhani, P., Asadi, S., Ghobakhloo, M., Dwivedi, Y. K., & Tseng, M.-L. (2023). Effects of supply chain transparency, alignment, adaptability, and agility on blockchain adoption in supply chain among SMEs. *Computers & Industrial Engineering*, 176, 108931. <https://doi.org/10.1016/j.cie.2023.108931>
- [27] Kairouz, P., McMahan, H. B., Avent, B., et al. (2021). Advances and open problems in federated learning. *Foundations and Trends in Machine Learning*, 14(1–2), 1–210. <https://doi.org/10.1561/22000000083>
- [28] Kam, H.-J., Kim, D. J., & He, W. (2022). Should we wear a velvet glove to enforce information security policies in higher education? *Behaviour & Information Technology*, 41(10), 2259–2273. <https://doi.org/10.1080/0144929X.2021.1917659>
- [29] Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, 3, 422–440. <https://doi.org/10.1038/s42254-021-00314-5>
- [30] Konečný, J., McMahan, H. B., Yu, F. X., et al. (2016). Federated learning: Strategies for improving communication efficiency. *arXiv*. <https://doi.org/10.48550/arXiv.1610.05492>

- [31] Li, H., Wu, H., and collaborators. (2024). Federated learning data security and privacy-preserving in IoT: A review. *Artificial Intelligence Review*, 57, 1–44. <https://doi.org/10.1007/s10462-024-10774-7>
- [32] Lumineau, F., Wang, W., & Schilke, O. (2021). Blockchain governance—A new way of organizing collaborations? *Organization Science*, 32(2), 500–521. <https://doi.org/10.1287/orsc.2020.1379>
- [33] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*. <https://doi.org/10.48550/arXiv.1705.07874>
- [34] March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, 2(1), 71–87. <https://doi.org/10.1287/orsc.2.1.71>
- [35] McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. *arXiv*. <https://doi.org/10.48550/arXiv.1602.05629>
- [36] Mnih, V., Kavukcuoglu, K., Silver, D., et al. (2015). Human-level control through deep reinforcement learning. *Nature*, 518, 529–533. <https://doi.org/10.1038/nature14236>
- [37] Ocasio, W. (1997). Towards an attention-based view of the firm. *Organization Science*, 8(3), 187–206. <https://doi.org/10.1287/orsc.8.3.187>
- [38] Pikhart, M., & Al-Obaydi, L. H. (2025). Reporting the potential risk of using AI in higher education: Subjective perspectives of educators. *Computers in Human Behavior Reports*, 18, 100693. <https://doi.org/10.1016/j.chbr.2025.100693>
- [39] Poppo, L., Zhou, K. Z., & Li, J. J. (2016). When can you trust “trust”? Calculative trust, relational trust, and supplier performance. *Strategic Management Journal*, 37(4), 724–741. <https://doi.org/10.1002/smj.2369>
- [40] Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks. *Journal of Computational Physics*, 378, 686–707. <https://doi.org/10.1016/j.jcp.2018.10.045>
- [41] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?” Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. <https://doi.org/10.1145/2939672.2939778>
- [42] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *arXiv*. <https://doi.org/10.48550/arXiv.1505.04597>
- [43] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv*. <https://doi.org/10.48550/arXiv.1707.06347>
- [44] Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. *Communications of the ACM*, 63(12), 54–63. <https://doi.org/10.1145/3381831>
- [45] Sirmon, D. G., Hitt, M. A., Ireland, R. D., & Gilbert, B. A. (2011). Resource orchestration to create competitive advantage. *Journal of Management*, 37(5), 1390–1412. <https://doi.org/10.1177/0149206310381869>
- [46] Staw, B. M., Sandelands, L. E., & Dutton, J. E. (1981). Threat-rigidity effects in organizational behavior. *Administrative Science Quarterly*, 26(4), 501–524. <https://doi.org/10.2307/2392337>
- [47] Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, 18(7), 509–533. [https://doi.org/10.1002/\(SICI\)1097-0266\(199708\)18:7<509::AID-SMJ882>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1097-0266(199708)18:7<509::AID-SMJ882>3.0.CO;2-Z)
- [48] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. *arXiv*. <https://doi.org/10.48550/arXiv.1706.03762>
- [49] Verdecchia, R., Sallou, J., & Cruz, L. (2023). A systematic review of green AI. *WIREs Data Mining and Knowledge Discovery*, 13(4), e1507. <https://doi.org/10.1002/widm.1507>
- [50] Xue, Y., and collaborators. (2025). A comparative analysis of AI privacy concerns in higher education news coverage in China and Western countries. *Education Sciences*, 15(6), 650. <https://doi.org/10.3390/educsci15060650>

- [51] Yang, Z., and collaborators. (2026). Comparing the sustainable role of higher education in artificial intelligence governance. *Sustainability*, 18(8), 3831. <https://doi.org/10.3390/su18083831>
- [52] Zhao, J., and collaborators. (2024). Joint client and resource optimization for federated learning systems. *Applied Sciences*, 14(2), 542. <https://doi.org/10.3390/app14020542>
- [53] Zhu, X., and collaborators. (2024). Federated learning-based IoT: A systematic literature review. *International Journal of Communication Systems*, 35(9), e5185. <https://doi.org/10.1002/dac.5185>