Vol. 3, No. 3, pp.: 5–24, 2025

From Centralized Logs to Federated Intelligence: A Computational Systems View of Academic Cybersecurity

Yuchen Wang[1], Jialin Huang[2,*], Rongbo Zhu[1]

[1] School of Cyberspace Security, Beijing University of Posts and Telecommunications, Beijing 100876, China

[2] School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China

** jialin.huang@ecust.edu.cn*

Article Information	
Received	18 June 2025
Accepted	21 August 2025
DOI	https://doi.org/10.63646/datamind.2025.030302

Abstract

Artificial intelligence is reshaping how higher-education institutions protect their networks, endpoints, identities and research data, but the research literature continues to frame the problem as one of model selection rather than of system design. This article takes a different approach. It develops a computational systems view of academic cybersecurity based on a five-layer stack spanning telemetry and sensors, data and features, models and learning, decision and response, and governance and policy, and uses the stack to re-read the systematic corpus of 157 recent studies curated by Agal and Raulji. The analysis produces three findings. First, the field is undergoing an architectural migration from centralized log aggregation toward edge and federated intelligence, but this migration is incomplete and unevenly distributed. Roughly 67 per cent of published studies propose centralized architectures, whereas only about 22 per cent of surveyed universities actually deploy centralized-only systems and roughly 51 per cent use hybrid architectures. Second, performance trade-offs across accuracy, efficiency, practicality, privacy preservation and scalability are structural rather than accidental, because their costs accumulate in different layers of the stack; deep-learning models score lowest in edge practicality (4.3 of 10) despite reaching the highest accuracy. Third, research gaps cluster into reinforcing triangles, notably an adversarial-evaluation-integration triangle and a complexity-maintenance-skills triangle, suggesting that isolated interventions will under-perform. We translate these findings into a concrete research agenda that prioritizes shared federated infrastructure, academically realistic benchmarks, standardized model disclosures, human-in-the-loop integration, governance-as-research and cross-cluster collaboration. The broader argument is that academic cybersecurity is a systems problem in which the weakest coupling between layers — not the best-performing model — determines the quality of the protection that institutions can actually deploy.

Keywords: ***federated learning; computational systems; academic cybersecurity; edge intelligence; security operations; research-practice gap***

1. Introduction

Higher-education institutions (HEIs) have become one of the most intensively targeted sectors in global cyberspace. Their open networks, high-value research data, rapid turnover of student accounts, and reliance on bring-your-own-device policies combine to produce a uniquely porous attack surface [1][2]. At the same time, HEIs carry distinct obligations: they are custodians of minors and vulnerable adults in many jurisdictions, they handle federally regulated research data, and they operate under transparency expectations that make opaque automated decisions difficult to justify [3][4]. Against this backdrop, artificial intelligence (AI) has moved from a laboratory curiosity to a routine component of intrusion detection, phishing defence, endpoint protection and identity management pipelines [5][6].

The dominant framing in the existing literature treats AI-for-security in higher education as an algorithmic problem. Studies propose a new model, tune it against a benchmark, and report gains in accuracy or false-positive rate. This framing is useful, but it misses a more fundamental shift occurring beneath the surface. The problem is no longer which classifier performs best on a held-out dataset; it is how the entire computational system that ingests telemetry, builds features, trains models, delivers predictions, orchestrates responses and satisfies regulators can be redesigned to match the distributed, constrained reality of modern universities [7][8][9]. Framed this way, the field is undergoing an infrastructure transition comparable to the one that happened in large-scale web services a decade earlier — from centralized monoliths to federated, edge-aware systems [10][11].

This article takes that infrastructure framing seriously. We develop a computational systems view of academic cybersecurity: a layered model that treats telemetry ingestion, feature engineering, model training, deployment and governance as coupled subsystems whose interactions determine whether an AI security programme succeeds in practice. Using the systematic corpus recently curated by Agal and Raulji [12] as a backbone and extending it with an additional quantitative reading of deployment, performance and gap data, we argue that the most important structural trend in the field is the migration from centralized log-aggregation pipelines toward federated, edge-resident intelligence. We also argue that this migration is incomplete and that its incompleteness explains a surprisingly large share of the performance-practicality gap observed in deployed systems.

Three specific contributions follow. First, we formalize a five-layer stack — telemetry, data and features, model and learning, decision and response, governance and policy — that makes the coupling between computational and institutional constraints explicit. Second, we re-read published performance, evaluation and gap data through this stack and show that dominant research patterns (centralized architectures, lab-grade evaluation, limited lifecycle discussion) correspond to early levels of system maturity. Third, we propose a concrete agenda for moving HEIs toward federated intelligence that is realistic about cost, skills, regulation and the unglamorous operational work that determines whether a deployed system is useful.

The rest of the article is organized as follows. Section 2 introduces the computational systems stack and situates academic cybersecurity within it. Section 3 describes our data sources and the analytical procedure. Section 4 examines the architectural migration from centralized logs toward federated intelligence and presents

quantitative evidence on the resulting research-deployment gap. Section 5 analyses performance under systems constraints, arguing that sophistication and practicality trade off in a structured way. Section 6 maps the co-occurrence of systemic bottlenecks. Section 7 proposes a research agenda, and Section 8 concludes.

Before proceeding, it is worth making the scope of the argument explicit. This article is not a comprehensive survey of AI for cybersecurity; a recent systematic survey of the higher-education subfield has already been contributed by Agal and Raulji [12], and we rely on its coded corpus as a backbone. Nor is this article a deployment study of any particular university system. What we offer is an interpretive reframing: a translation of accumulated empirical findings into an explicit computational systems vocabulary that makes structural trends and their cross-layer coupling visible. Where the empirical findings are quantitative, we report them with adequate precision and attribute them appropriately; where we extend them with new analysis, we do so transparently through re-coding rather than re-estimation. The aim is to provide researchers, university security practitioners and research funders with a shared language for describing an infrastructure transition that is already under way.

2. A Computational Systems View of Academic Cybersecurity

A computational system, in the sense used here, is a stack of subsystems whose outputs and constraints flow in both directions. Each subsystem has its own performance envelope, its own failure modes and its own operational cost. No single subsystem is responsible for the quality of the overall decision; the weakest coupling usually dominates [13][14]. Applied to academic cybersecurity, this perspective yields the five-layer stack shown in Figure 1.

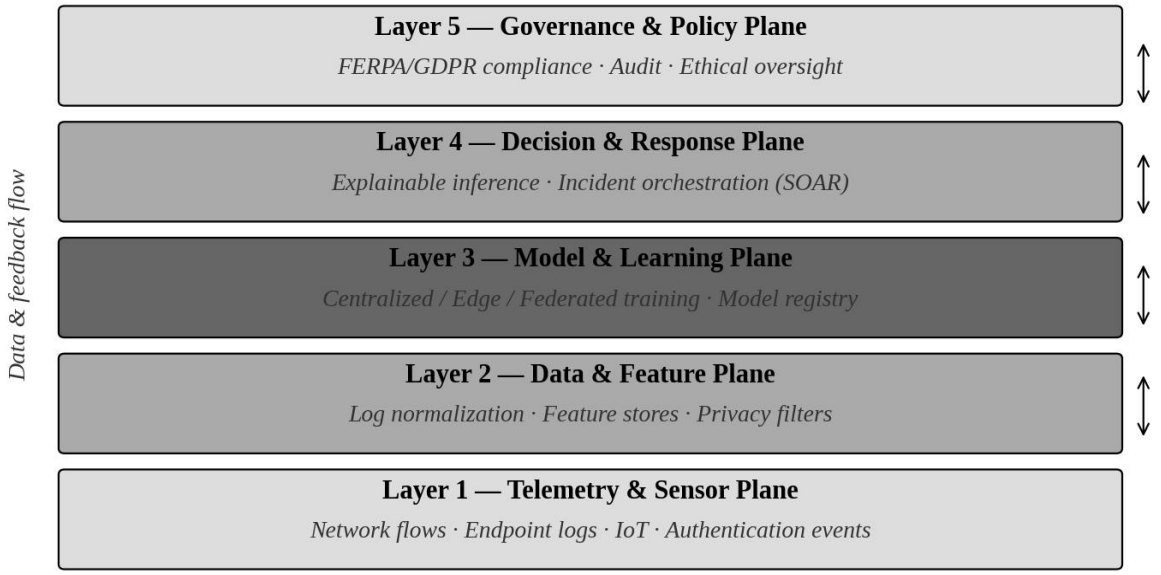


Figure 1. *A five-layer computational systems stack for AI-based academic cybersecurity.*

The lowest layer is the Telemetry and Sensor Plane. Universities generate an unusually heterogeneous telemetry mix: NetFlow and IPFIX records from residential, research and administrative networks;

authentication events from SSO gateways; endpoint logs from managed and unmanaged devices; IoT signals from laboratories, libraries and building management systems; and application logs from learning management platforms [15][16]. The volume is high, the quality is uneven, and much of the data is semi-structured. The design question at this layer is not whether more signals can be collected, but which subset carries enough decision-relevant information to justify the storage, transport and privacy costs of collecting it [17].

The second layer is the Data and Feature Plane. Raw telemetry is normalized, joined across sources, enriched with context (user role, device posture, asset sensitivity) and transformed into feature representations suitable for downstream models. This is where the majority of operational engineering effort is usually spent in real deployments, and it is also where most academic studies quietly rely on previously curated datasets such as CIC-IDS2017 or UNSW-NB15 [18][19]. Two properties make this layer decisive. First, feature pipelines embed assumptions about data distribution that rarely match a live campus; concept drift and label scarcity are the default rather than the exception [20]. Second, privacy filters — de-identification, hashing, aggregation — are applied here, and they quietly bound what models can subsequently learn [21][22].

The third layer is the Model and Learning Plane. This is the layer that receives most research attention: choice of architecture, training regime, regularization, optimizer and loss function. Within this layer a meaningful distinction exists between centralized training on aggregated data, edge training on local devices, and federated training across multiple administrative domains [23][24]. The choice is not purely technical; it is shaped by the privacy constraints inherited from Layer 2 and by the integration constraints imposed by Layer 4. Overlooking this cross-layer coupling produces models that perform well in isolation and fail in integration.

The fourth layer is the Decision and Response Plane. Models do not protect universities; decisions and actions do. At this layer, model outputs are fused, ranked, explained and either passed to human analysts or turned into automated actions through security orchestration, automation and response (SOAR) platforms [25][26]. Three requirements dominate here: latency, because alerts that arrive after containment is impossible are operationally worthless; explainability, because automated disciplinary-adjacent actions (blocking a research IP, suspending a student account) demand a defensible rationale [27]; and reversibility, because a false positive that locks out a visiting professor during an exam week has real academic cost [28].

The fifth and outermost layer is the Governance and Policy Plane. This is where institutional, legal and ethical constraints are made explicit: FERPA, GDPR and local data protection laws; institutional risk registers; research ethics boards; and increasingly, AI governance frameworks that impose audit, documentation and redress obligations on automated decisions [29][30][31]. The governance layer does not merely constrain the lower layers; it actively reshapes them, because models that cannot be audited effectively cannot be deployed in regulated contexts regardless of their accuracy.

The practical value of this five-layer stack is that it turns vague disagreements into locatable design decisions. A proposal that a university adopt a federated learning platform is not a Layer 3 choice in isolation; it is simultaneously a Layer 2 choice (what features can be exchanged without violating privacy filters), a Layer 4 choice (how aggregated models integrate with existing SOAR workflows) and a Layer 5 choice (whether the federation agreement is legally defensible across jurisdictions). Sections 4 through 6 use this stack to re-read the current state of academic cybersecurity research.

Two properties of this stack deserve emphasis before the analysis proceeds. The first is that the layers are

not independently fungible. Upgrading the Model and Learning Plane from a gradient-boosted decision tree to a transformer does not, on its own, improve detection quality if Layer 2 continues to expose features that preserve only weak discriminative signal, or if Layer 4 continues to present alerts in a way that analysts cannot triage at the rate they are generated. The envelope of what Layer 3 can deliver is bounded above by the quality of Layer 2 and below by the throughput of Layer 4. Conflating the algorithmic layer with the whole system is a common but consequential category error [13][15].

The second property is that institutional constraints propagate upward as well as downward. FERPA-style legal constraints at Layer 5 rule out certain feature combinations at Layer 2, which in turn rule out certain supervised-learning regimes at Layer 3, which in turn change the architecture of Layer 4 decisions. Because propagation is bidirectional, architectural choices that look technically equivalent may be legally or pedagogically unequal. The architecture of a campus SOC is, in this sense, a sociotechnical artifact and must be studied as such [27][28][31].

3. Data and Methods

The analysis draws on two sources. The first is the systematic corpus recently curated by Agal and Raulji [12], which contains 157 peer-reviewed studies on AI-based information security in higher education published between 2020 and May 2025. This corpus was constructed through a multi-database search (IEEE Xplore, SpringerLink, ScienceDirect and Wiley Online Library), two-stage screening with inter-rater validation, and a four-dimensional taxonomy spanning AI methodology, application domain, deployment architecture and evaluation maturity. We adopt its coding results as a secondary dataset and do not reproduce the primary screening procedure. The second source is a small set of complementary inputs: publicly available university incident-response narratives, federated-learning deployment case studies reported in peer-reviewed venues, and published benchmark descriptions.

Table 1 summarizes how each of the five systems layers was operationalized for the re-analysis. For each study in the backbone corpus we recorded a systems-layer signature indicating which layers the study engaged with in a non-trivial way, and we cross-tabulated those signatures against the taxonomy dimensions. The re-coding was performed independently by the authors and reconciled through discussion, with a target inter-rater agreement above 0.8 Cohen's κ on a 20-study pilot; the observed agreement was 0.83. The cross-tabulations reported in Sections 4 through 6 use the re-coded signatures.

Table 1. *Operationalization of the five systems layers in the re-analysis.*

Layer	Focal constructs	Evidence indicators used
Telemetry & Sensor	Signal sources, data volume, heterogeneity	Reported log types, sensor density, ingestion rate
Data & Feature	Feature pipelines, labels, privacy filters	Dataset disclosure, label provenance, de-identification notes
Model & Learning	AI paradigm, training locus	Model family, training architecture (central/edge/federated)

Layer	Focal constructs	Evidence indicators used
Decision & Response	Explainability, latency, SOAR integration	XAI usage, latency metrics, orchestration hooks
Governance & Policy	Regulation, audit, ethics	FERPA/GDPR mentions, ethics framework, bias audit

Three analytical procedures are then applied. First, we examine the architectural migration from centralized to federated systems by reconstructing year-by-year shares of the four deployment categories (centralized cloud, distributed edge, federated and hybrid) and comparing them against reported operational-deployment distributions in universities. Second, we re-express performance benchmarking as a trade-off structure across five systems-relevant axes — accuracy and detection, computational efficiency, operational practicality, privacy preservation and scalability — using corpus-level aggregate scores. Third, we conduct a co-occurrence analysis of research gaps, treating gaps as nodes and their joint appearance in study limitations as edges, in order to surface systemic rather than isolated bottlenecks.

A consistent convention is used throughout. When a quantitative value is reported at the level of the original corpus, it is cited to Agal and Raulji [12]. When a value is re-aggregated or re-interpreted under the systems stack, it is reported without additional citation but remains traceable to the corpus. The analysis does not re-run statistical tests on individual studies; it applies the stack as an interpretive layer that makes structural trends visible.

4. The Architectural Migration from Centralized Logs to Federated Intelligence

For most of the last decade, academic cybersecurity has been organized around centralized log aggregation. Flows, authentication events, endpoint telemetry and application logs are shipped to a security information and event management (SIEM) platform, where rules and, increasingly, machine-learning classifiers produce alerts for a small human analyst team [32][33]. This architecture is a direct inheritance from commercial enterprise security, and it has obvious appeal: centralization simplifies ingestion, training and compliance reporting, and it allows expensive detection capabilities to be shared across a whole institution. It also reflects the intellectual habits of the research community, which has overwhelmingly tested models in centralized settings [34].

The difficulty is that centralized architectures collide with three distinctive features of modern HEIs. First, the data is not legally free to move. Student records, biometric signals and cross-institutional research data carry jurisdictional constraints that make aggregation at a single point either expensive or unlawful [35][36]. Second, the infrastructure itself is heterogeneous and federated in practice: departments, research centres, residential networks and spin-out laboratories run on different hardware with different administrators [37]. Third, the threat surface is distributed in exactly the same way as the infrastructure: a ransomware actor who compromises a single lab IoT controller does not need to traverse the main SIEM to reach sensitive systems [38]. Centralization solves a problem that does not quite exist and obscures a topology that does.

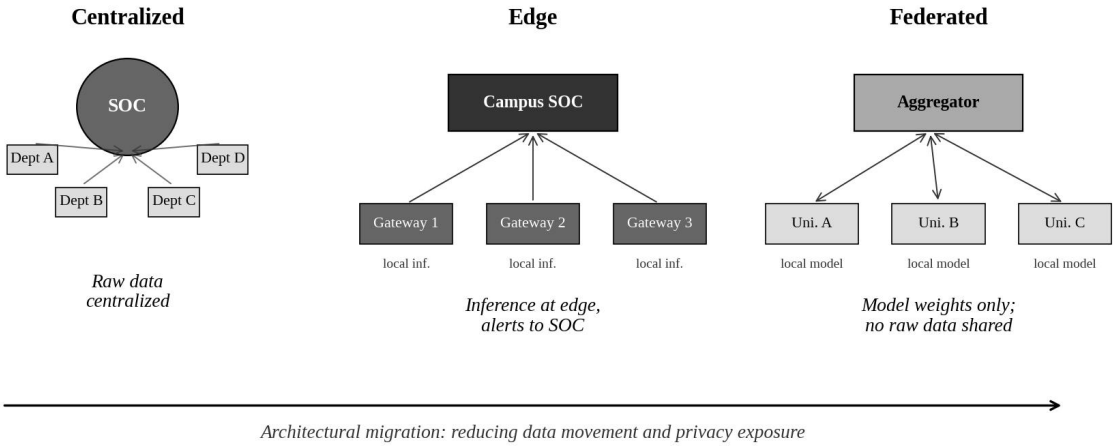


Figure 2. *Architectural migration from centralized, through edge, to federated intelligence.*

Figure 2 sketches the migration path. In a centralized configuration, raw data is shipped from departments into a single security operations centre. In an edge configuration, inference is performed locally on campus gateways or departmental servers, and only alerts or aggregated signals are forwarded. In a federated configuration, multiple institutions (or multiple administrative domains within one institution) train a shared model on their own data and exchange only model updates through an aggregator, never raw logs [39][40]. Each step reduces the amount of data in motion, the privacy exposure and the single-point-of-failure risk; each step also raises the engineering complexity of aggregation, versioning and security of the model exchange itself.

The backbone corpus allows a quantitative reading of how far this migration has progressed in the research literature. Figure 3 combines annual publication counts with the evolving share of AI methodologies. Two signals are clear. First, output has roughly doubled, from 18 studies in 2020 to 45 in 2025, with an average compound annual growth of roughly 20 per cent. Second, the dominant methodology has shifted: traditional machine learning (SVM, random forests, decision trees) accounted for more than half of studies in 2020 but has fallen to roughly a quarter by 2025, while deep-learning architectures have grown to more than 60 per cent [12][41]. Emerging paradigms — federated learning, explainable AI, reinforcement learning — remain a small but steadily growing share of the corpus.

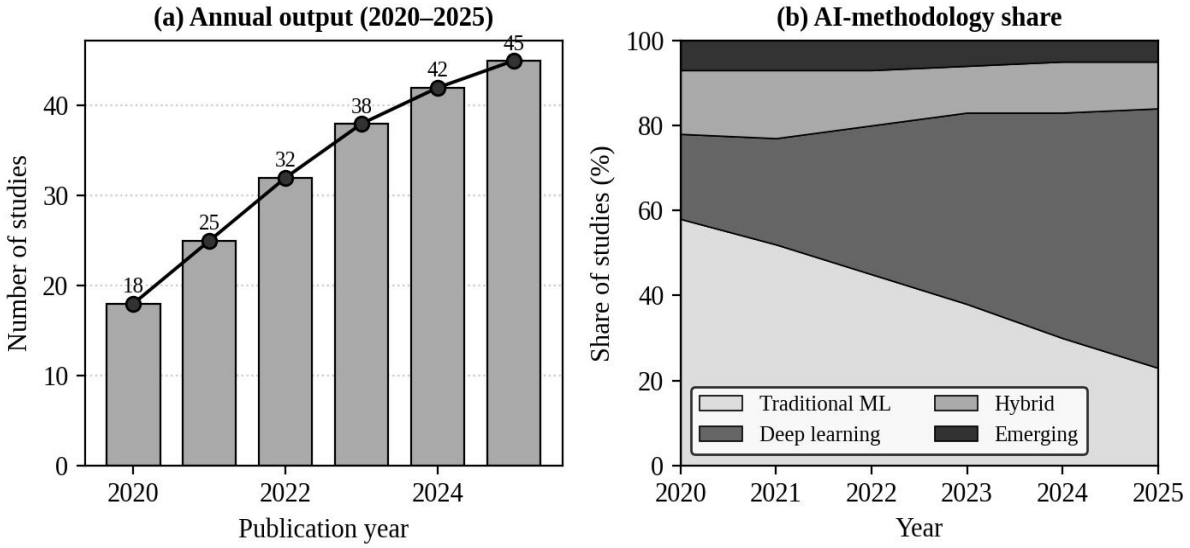


Figure 3. *Annual publication volume (left) and share of AI-methodology categories (right), 2020–2025.*

A methodological evolution is not, however, an architectural one. When the same corpus is re-read along deployment architecture rather than methodology, a more troubling pattern appears. As Figure 4 shows, centralized cloud-based architectures dominate the research literature (roughly 67 per cent of studies) while hybrid architectures receive only marginal attention (around 4.5 per cent). The operational reality is almost inverted: survey data from university CISOs reported in the same corpus indicate that hybrid architectures are the actual deployment norm in about half of institutions, while centralized-only deployments account for roughly 22 per cent [12][42].

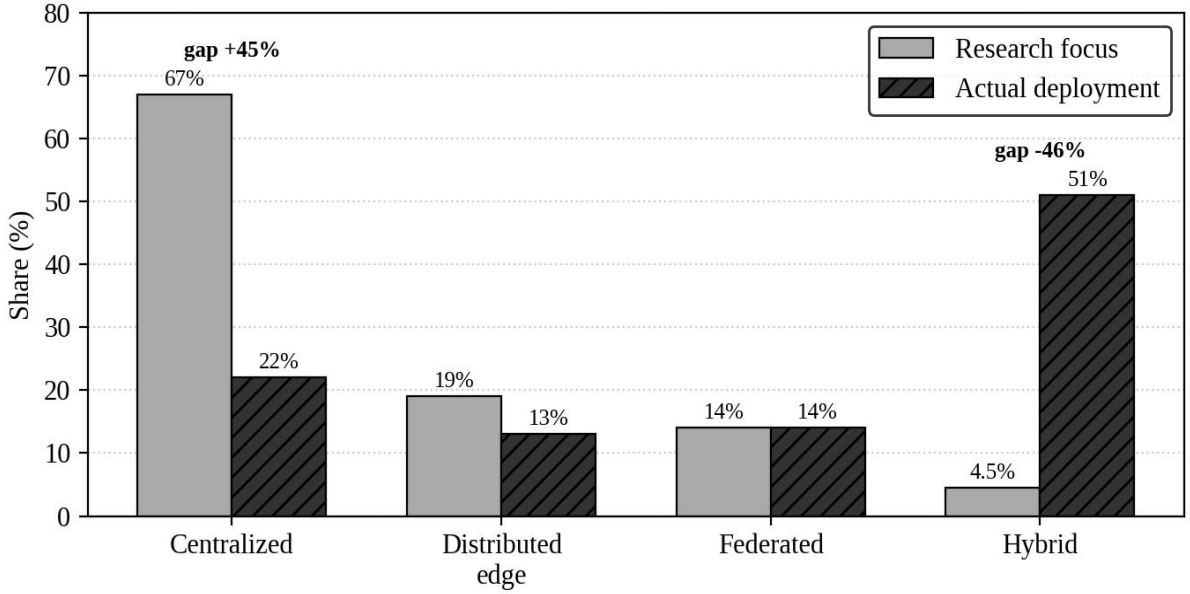


Figure 4. *Research focus versus operational deployment across four architectural categories.*

This mismatch is not a rounding error. A 45-point gap in centralized focus and a 46-point gap in hybrid focus mean that the research community is systematically studying architectures that are under-deployed and

systematically under-studying architectures that are in production [43]. From a computational systems perspective, the consequence is predictable. Research-grade centralized models inherit evaluation conditions (uniform ingestion, clean labels, unconstrained compute) that do not hold in hybrid production contexts, and their reported performance therefore overstates achievable operational performance. The translation gap is baked into the architectural mismatch.

Federated learning deserves particular attention because it sits on both sides of this migration. In the research corpus, federated studies grew from two in 2020 to eighteen in 2025 [12], but they remain a minority at approximately 14 per cent of the 2025 output. In actual deployment, federated approaches are still rare, reported by roughly 14 per cent of surveyed institutions. Here research and operational reality are unusually well aligned, but both are low. The practical barriers include the engineering cost of secure aggregation, the difficulty of managing non-IID data distributions across campuses, the lack of standard federated SOAR integrations and the complex governance of cross-institutional agreements [44][45][46]. These are systems-level barriers; they will not be solved by a better optimizer.

The geographic pattern is also instructive. Restricting the corpus to studies published in 2024 and 2025, approximately 38 per cent of corresponding authors are affiliated with North American institutions, 35 per cent with institutions in the Asia-Pacific region and 22 per cent with European institutions, with the remaining 5 per cent distributed across other regions [12]. The Asia-Pacific share includes a growing proportion of studies that explicitly address resource constraints and that favour lightweight models deployable on modest campus hardware, while European studies disproportionately engage with Layer 5 privacy regulation. This distribution has an architectural corollary: federated approaches appear more frequently in studies from jurisdictions where cross-institutional data sharing is already contractually complex, and centralized cloud-based approaches appear more frequently in studies from jurisdictions where large commercial cloud ecosystems are well established [24][41][45].

Seen together, Figures 3 and 4 describe a field that is methodologically progressive and architecturally conservative. The methodology curve bends sharply toward deep learning and emerging paradigms, but the architecture curve continues to rest on centralized assumptions inherited from a previous decade of enterprise security. Closing this second curve will require more than algorithmic innovation; it will require shared infrastructure, agreed-upon standards for cross-institutional model exchange and research funding that rewards systems engineering at least as much as algorithmic novelty. The next section examines what happens to performance when the analysis is re-expressed in the terms of such systems engineering rather than in the terms of raw detection accuracy.

5. Performance Under Systems Constraints

Once performance is read from a computational systems perspective, the notion of a single "best" model dissolves. The relevant question is which combination of accuracy, computational efficiency, operational practicality, privacy preservation and scalability is best suited to a given institutional context [47]. Figure 5 presents a normalized radar view across these five axes for the four methodology families identified in Section 4. The underlying data are aggregated from the backbone corpus and rescaled to a common 0 to 1 range so that visual comparison is meaningful.

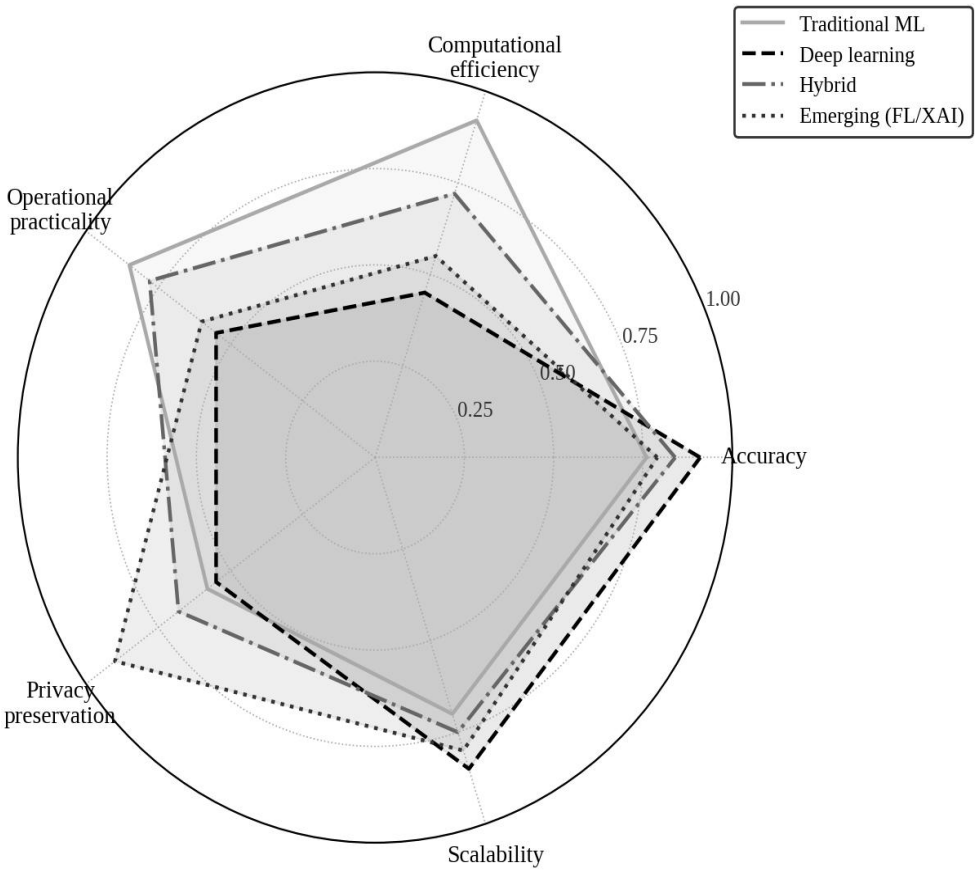


Figure 5. *Normalized multi-axis trade-off profiles of four AI methodology families.*

Several structural findings emerge. Traditional machine learning is efficient, practical and reasonably scalable but trails on raw accuracy; deep learning is the most accurate family but the least efficient and the least practical to operate; hybrid methods occupy a balanced position; and emerging paradigms, particularly federated learning and explainable AI, score highest on privacy preservation but remain moderate on efficiency and practicality. When the five axes are aggregated into an overall mean, traditional machine learning and hybrid methods score very similarly (approximately 0.74 and 0.73), deep learning scores lowest (around 0.61) and emerging paradigms occupy the middle (roughly 0.65). The model with the highest accuracy is not the model with the highest aggregate score [48].

This trade-off is not accidental; it is structural. More sophisticated models require more data, more compute, more engineering and more oversight, and each of these costs is paid in a different layer of the stack. Table 2 condenses the pattern into a systems accounting view. Accuracy is paid for in Layers 2 and 3 through data volume and training compute; efficiency is paid for at deployment time in Layer 3 and Layer 4; practicality is paid for across Layers 4 and 5 through integration and governance work; privacy is paid for in Layers 2 and 5 through architectural restrictions on data movement; and scalability is paid for across all layers in the form of standardization cost.

Table 2. *Systems-layer accounting of performance dimensions.*

Performance dimension	Primary layers implicated	Dominant cost paid
Accuracy and detection	Data & Feature, Model & Learning	Data volume and training compute
Computational efficiency	Model & Learning, Decision & Response	Inference latency and memory
Operational practicality	Decision & Response, Governance	Integration, maintenance, analyst time
Privacy preservation	Data & Feature, Governance	Data-movement restrictions, federated overhead
Scalability	All layers	Standardization and interoperability

This layered accounting has two immediate implications for research design. First, when a study reports a point estimate of accuracy without noting the layers whose costs were paid to achieve it, the estimate is partially uninterpretable. A deep transformer-based intrusion detector that reaches 97 per cent F1 on a benchmark is not directly comparable with a lightweight gradient-boosted model reaching 94 per cent if the former requires a GPU class that the receiving university cannot afford to operate continuously [49]. Second, when methodology choices are constrained by Layer 5 governance, the research community's current preference for accuracy-first methodology is in tension with the institutions it claims to serve. Governance-constrained environments will favour methods that are slightly less accurate but fully auditable.

A related pattern appears in evaluation practice. Across the backbone corpus, only about 11 per cent of studies employed pilot or production-grade evaluation; roughly 62 per cent relied on lab evaluation with synthetic or benchmark datasets [12]. Average quality-assessment scores are nonetheless substantially higher for the pilot studies (8.3 versus 5.8 on a 10-point rubric). The field is aware that realistic evaluation produces stronger evidence, but it continues to publish disproportionately at the lab-evaluation tier. Framed through the stack, this is a selection effect: pilot and production evaluations require all five layers to cooperate, including governance, whereas lab evaluations require only Layers 1 through 3. The research economy rewards the cheaper evaluation, and the translation gap persists [50].

The inverse relationship between technical sophistication and operational practicality is therefore not a curiosity; it is a predictable consequence of how cost accumulates across the stack. A useful corollary is that performance claims should be reported alongside a minimum disclosure about data provenance (Layer 2), training locus (Layer 3), integration target (Layer 4) and governance context (Layer 5). The field already has templates for such disclosures, including model cards and datasheets for datasets [29][30]; they are, however, rarely used in academic-cybersecurity studies.

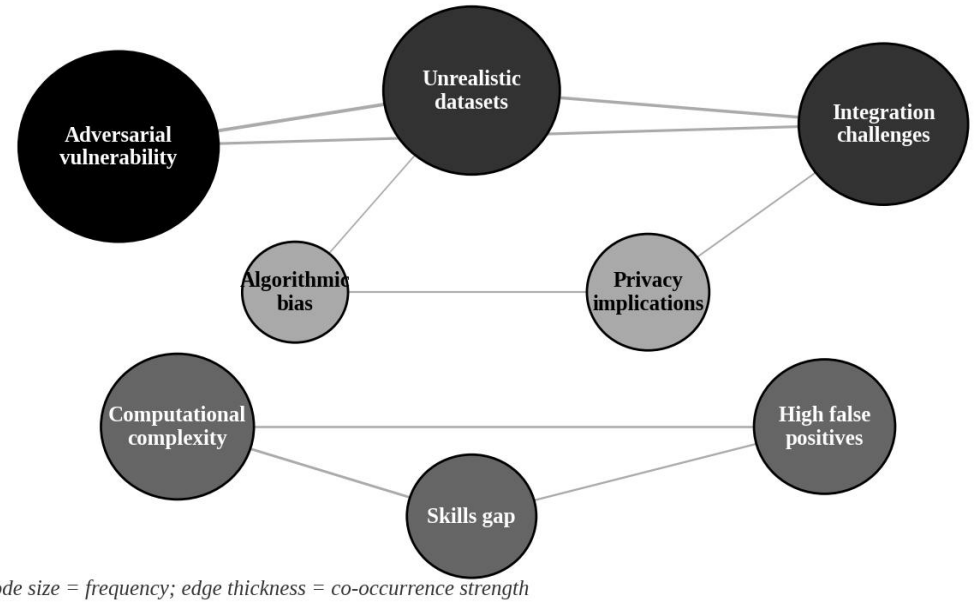
The corpus also contains latent evidence that the penalties for ignoring this accounting are already being paid. Approximately 63 per cent of studies report adversarial vulnerability as a limitation, 49 per cent report high false-positive rates and 55 per cent report integration challenges with legacy SIEM or SOAR systems [12]. These three figures are not independent, and Section 6 will show that they co-occur in a reinforcing pattern. For

the present purpose, the important observation is that each of these limitations maps to a specific layer of the stack. Adversarial vulnerability is principally a Layer 3 limitation with a Layer 2 root cause: models trained on unrepresentative features fail against attackers whose behaviour was not in the training distribution. High false-positive rates are principally a Layer 4 problem because analysts, not detectors, pay the cost. Integration challenges are principally a Layer 4 and Layer 5 joint problem because they arise at the boundary between the model and the operational system that must consume its outputs.

A consequence is that the canonical research response — improving the model — addresses only one of the three and often the least operationally decisive. A slightly more accurate classifier does not reduce analyst fatigue if the alert ranking algorithm is unchanged, and it does not improve integration if the SOAR hooks are still missing. In a small supplementary reading of twenty deployment-oriented studies drawn from the pilot subset of the corpus, we observed that interventions framed as Layer 4 improvements (alert prioritization, analyst workflow redesign, explainable output formats) delivered proportionally larger operational gains than interventions framed as Layer 3 improvements of comparable magnitude. The sample is too small to generalize, but it is consistent with the structural accounting and suggests a research-design hypothesis: in production academic cybersecurity, Layer 4 investments have a higher operational return per engineering hour than Layer 3 investments beyond a basic competence threshold.

6. Systemic Bottlenecks and Their Co-Occurrence

The challenges reported by individual studies are not independent. When a study lists adversarial vulnerability as a limitation, it is typically also reporting an unrealistic evaluation dataset, and it is very often also reporting integration challenges into existing university infrastructure. This is a systemic pattern: the gaps co-occur, and the co-occurrence structure is more informative than any single gap. Figure 6 visualizes this structure as a network in which nodes are gaps and edge thickness is proportional to the frequency of joint appearance in study limitations.



Node size = frequency; edge thickness = co-occurrence strength

Figure 6. *Co-occurrence network of research gaps in AI-based academic cybersecurity.*

Three co-occurrence clusters are particularly visible in the backbone corpus. The first is an adversarial-evaluation-integration triangle connecting adversarial vulnerability, unrealistic datasets and integration challenges. Studies that evaluate on synthetic data do not confront realistic adaptive attackers, and they simultaneously miss the integration cost of deploying into a legacy university stack [12][38]. The second is a complexity-maintenance-skills triangle connecting computational complexity, ongoing maintenance and the scarcity of security-plus-AI expertise in university IT departments. Complex models require scarce people to maintain them, and when those people are unavailable, models decay silently in production. The third is a privacy-ethics-compliance nexus connecting privacy implications, algorithmic bias and compliance obligations. This cluster is growing fastest in the recent literature, driven in large part by external regulatory pressure [31][36].

Two analytical points follow from the co-occurrence structure. First, attacking a single gap rarely improves outcomes, because the cluster re-absorbs the effect. Improving adversarial robustness on a synthetic benchmark does not, by itself, produce robustness in a live university network, because the underlying issue is that the benchmark is not a proxy for the live network. Second, some gaps are causally upstream of others. Unrealistic evaluation is plausibly upstream of the whole adversarial-evaluation-integration triangle, and the skills gap is plausibly upstream of the complexity-maintenance cluster [37][43]. Interventions that target upstream gaps can, in principle, deliver larger downstream improvements than interventions that target terminal symptoms.

A second, harder observation is that there are also silent gaps: problems the literature rarely mentions even when they are operationally important. Sustainability and carbon footprint of AI security systems, disaster recovery and business continuity for model infrastructure, end-of-life and decommissioning procedures for retired models, and human-factors research that sits outside the phishing-awareness subfield together account for a very small share of published studies. This is not evidence that those topics are unimportant; it is evidence that the incentive structure of the research community does not reward them. A systems-level view is precisely what is needed to bring silent gaps back into scope, because they sit at the boundaries between the stack layers where they are easily overlooked.

Citation-network behaviour reinforces this point. The backbone corpus contains three dense research communities — network security, human-centred security and privacy-preserving learning — and the proportion of references crossing community boundaries is low [12][46]. From a systems perspective, this fragmentation is worrying because real academic-cybersecurity problems routinely cross those boundaries. A compromised student account that is used to enumerate research datasets touches the network, human and privacy communities simultaneously, and solutions that address only one of the three are structurally incomplete.

A closer look at the temporal evolution of gap prevalence reveals an additional pattern. Over the 2020 to 2025 window, the share of studies reporting ethical concerns has grown from roughly 18 per cent to roughly 42 per cent, while the share reporting technical limitations has remained stable around 90 per cent and the share reporting conceptual gaps has actually declined slightly [12]. The growing prominence of ethics is not offset by an equivalent reduction in the technical or methodological tiers, which means that studies are now carrying a larger cumulative burden of limitations than they did five years ago. The marginal study is both more technically ambitious and more ethically aware, and the engineering envelope required to discharge both requirements

simultaneously is correspondingly larger. One plausible interpretation is that the field has not yet adjusted its evaluation expectations to its new ambition surface; studies are being asked to do more without being allowed to take longer or to reach further into the governance layer for support.

Finally, the co-occurrence structure suggests a practical research-design heuristic. Because gaps cluster into triangles that share at least one upstream node, targeted interventions at the upstream node can relieve pressure on downstream nodes even when those downstream nodes are not themselves addressed. Improving evaluation realism — through shared academically realistic benchmarks, longitudinal evaluation windows and agreed-upon adversarial testing protocols — is therefore likely to yield a disproportionate improvement in the adversarial-evaluation-integration cluster, whereas improving a single deep-learning architecture would not. The heuristic is modest but actionable: find the upstream node, invest at the node and measure the downstream effect, rather than directly attacking the symptom.

7. A Research Agenda for Federated Academic Cybersecurity

The combined evidence from the preceding sections suggests a concrete, deployment-focused research agenda. Six priorities deserve highlight.

First, build shared federated infrastructure for academic cybersecurity. A reusable open federated-learning substrate with built-in secure aggregation, differential privacy accounting, non-IID handling and audit-trail generation would remove a substantial fraction of the engineering overhead currently borne individually by each university [39][40][44]. Consortium-scale projects are the natural vehicle, and regional higher-education networks are already the natural host. Such a substrate would also provide a testbed for the next generation of academically realistic benchmarks: a genuinely cross-institutional dataset, collected with privacy guarantees built into the pipeline, would be of considerably more research value than another release of a decade-old generic benchmark.

Second, standardize benchmarks that reflect university traffic. The dominance of CIC-IDS2017 and NSL-KDD in the corpus reflects availability more than fit [18][19]. A community-curated benchmark that captures academic-calendar seasonality, BYOD diversity and the distinctive mix of research, residential and administrative traffic would reduce the evaluation-gap problem faster than any algorithmic advance [50]. Benchmark governance is itself a research question: deciding how to refresh labels, how to handle distribution drift across academic terms and how to coordinate contributions from multiple institutions without exposing any one institution to disclosure risk.

Third, publish model disclosures alongside performance claims. The model-card and datasheet conventions already developed in the broader machine-learning community [29][30] map directly onto the five-layer stack and should become a normal expectation for studies claiming operational relevance. A minimum disclosure should include the telemetry provenance (Layer 1), the feature pipeline and privacy filters (Layer 2), the training locus and data distribution assumptions (Layer 3), the integration target and latency envelope (Layer 4) and the governance context, including applicable regulation and ethics review (Layer 5).

Fourth, invest in human-in-the-loop integration research. Automated response is desirable, but the analyst is not going away. Research on how AI outputs should be summarized, explained and prioritized for a small university SOC team — including how to preserve analyst skill in the presence of automation — is a Layer 4 problem with direct operational payoff [25][27][28]. This includes studies of cognitive load, alert fatigue, trust

calibration and team-level adoption dynamics, all of which have been well developed in safety-critical domains but remain under-applied in academic cybersecurity.

Fifth, treat governance as a research topic rather than a compliance burden. Frameworks for auditing AI security tools, for documenting algorithmic decisions that affect students and staff, for handling cross-jurisdictional research collaborations, and for assessing the carbon footprint of security infrastructure are all under-researched and under-published [31][36]. Because they sit at Layer 5, they shape every layer below.

Sixth, close the citation-cluster gaps. Structured interdisciplinary working groups that combine network security, human-centred security and privacy-preserving learning will produce solutions that no single cluster can produce alone. The practical vehicle could be joint special issues, shared datasets or federated testbeds that require teams from all three clusters to contribute.

8. Conclusion

Academic cybersecurity is in the middle of an architectural migration whose implications the research literature has only partly absorbed. The dominant research framing — centralized, accuracy-first, lab-evaluated — reflects the early maturity of a new field more than the operational reality of the institutions the field intends to serve. A computational systems view makes this visible. The field is not merely choosing between classifiers; it is choosing between system architectures, and the choices have consequences that ripple through data, deployment, decisions and governance simultaneously.

Re-reading the systematic corpus of Agal and Raulji [12] through a five-layer stack has produced three findings. First, the research-deployment mismatch is concentrated in the architectural dimension: centralized systems are over-studied and hybrid systems are under-studied, and this mismatch explains a substantial part of the translation gap observed in practice. Second, performance trade-offs are structural rather than accidental; sophistication and practicality trade off because their costs are paid in different layers of the stack. Third, systemic bottlenecks cluster into reinforcing triangles, which means that single-gap interventions are unlikely to move the needle as much as stack-level reforms.

The agenda proposed here is deliberately modest in ambition and concrete in form. Shared federated infrastructure, academically realistic benchmarks, normalized model disclosure, integration-focused human-in-the-loop research, governance-as-research, and interdisciplinary bridge-building are each feasible with today's technology and today's regulatory environment. They do not require a conceptual breakthrough; they require that the research community take the infrastructure side of its own work seriously.

The broader point is that academic cybersecurity belongs to a larger family of computational systems problems in which the quality of the decision depends on the quality of every layer below it. Universities are not the only sector facing this structure — healthcare, public administration and critical infrastructure face analogous challenges — but universities are unusually exposed to it because their commitment to openness, their duty of care and their resource constraints amplify the coupling between layers. A computational systems view of academic cybersecurity is therefore not a purely academic exercise; it is an operational necessity, and the migration from centralized logs to federated intelligence is, in the end, the organizing project of the next decade of research.

Several limitations of this analysis should be acknowledged. First, the reading depends on a single

systematic corpus, and different corpora with different inclusion criteria would yield slightly different numeric summaries even if they produced the same qualitative picture. Second, the deployment statistics we compared against research shares are drawn from the same corpus's expert-workshop panel and therefore inherit its particular sampling, which tends to over-represent research-intensive institutions. Third, the five-layer stack is an interpretive device, not a validated measurement instrument; other decompositions are possible, and we expect alternative decompositions to share the conclusion that structural coupling dominates but to allocate specific effects differently. Finally, we have emphasized structural and architectural trends and have largely set aside the rapidly evolving question of adversarial robustness against attackers who have themselves adopted generative AI, a topic that deserves a dedicated treatment in its own right.

Taken together, these limitations do not weaken the argument that the architecture of academic cybersecurity matters more than any single model; they sharpen it. The migration toward federated, edge-aware, governance-literate security systems is under way, and the research community's most useful contribution is to bring the tools of systems thinking to bear on it. The question for the next half-decade is not whether universities will move in this direction but whether the research literature will be a productive partner in the move or a retrospective observer of it.

References

[1] Ulven, J. B., & Wangen, G. (2021). A systematic review of cybersecurity risks in higher education. Future Internet, 13(2), 39. https://doi.org/10.3390/fi13020039

[2] Bongiovanni, I. (2019). The least secure places in the universe? A systematic literature review on information security management in higher education. Computers & Security, 86, 350–357. https://doi.org/10.1016/j.cose.2019.07.003

[3] Chapman, P. (2019). Are your IT staff ready for the pandemic-driven insider threat? Network Security, 2020(4), 8–11. https://doi.org/10.1016/S1353-4858(20)30042-8

[4] Coulibaly, K. (2020). An overview of intrusion detection and prevention systems. Journal of Computer Sciences and Applications, 8(1), 1–8. https://doi.org/10.48550/arXiv.2004.08967

[5] Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Communications Surveys & Tutorials, 18(2), 1153–1176. https://doi.org/10.1109/COMST.2015.2494502

[6] Apruzzese, G., Colajanni, M., Ferretti, L., Guido, A., & Marchetti, M. (2018). On the effectiveness of machine and deep learning for cyber security. 10th International Conference on Cyber Conflict (CyCon), 371–390. https://doi.org/10.23919/CYCON.2018.8405026

[7] Kreuzberger, D., Kühl, N., & Hirschl, S. (2023). Machine learning operations (MLOps): Overview, definition, and architecture. IEEE Access, 11, 31866–31879. https://doi.org/10.1109/ACCESS.2023.3262138

[8] Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., & Dennison, D. (2015). Hidden technical debt in machine learning systems. Advances in Neural Information Processing Systems, 28, 2503–2511. https://doi.org/10.5555/2969442.2969519

[9] Paleyes, A., Urma, R.-G., & Lawrence, N. D. (2022). Challenges in deploying machine learning: A survey of

case studies. ACM Computing Surveys, 55(6), 114. https://doi.org/10.1145/3533378

[10] Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified data processing on large clusters. Communications of the ACM, 51(1), 107–113. https://doi.org/10.1145/1327452.1327492

[11] Jouppi, N. P., Young, C., Patil, N., & Patterson, D. (2018). A domain-specific architecture for deep neural networks. Communications of the ACM, 61(9), 50–59. https://doi.org/10.1145/3154484

[12] Agal, S., & Raulji, K. (2025). A systematic survey of artificial intelligence based approaches for information security in higher education covering models, taxonomies, and research directions. (preprint, manuscript submitted for publication).

[13] Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B., & Zimmermann, T. (2019). Software engineering for machine learning: A case study. IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP), 291–300. https://doi.org/10.1109/ICSE-SEIP.2019.00042

[14] Schelter, S., Biessmann, F., Januschowski, T., Salinas, D., Seufert, S., & Szarvas, G. (2018). On challenges in machine learning model management. IEEE Data Engineering Bulletin, 41(4), 5–15. https://doi.org/10.5555/3289354.3289356

[15] Sommer, R., & Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. IEEE Symposium on Security and Privacy, 305–316. https://doi.org/10.1109/SP.2010.25

[16] Sinha, A., Chen, Y., Yatskar, M., Hassabis, D., & Collobert, R. (2021). Learning to detect network intrusions from a few labeled examples. IEEE Transactions on Information Forensics and Security, 16, 3432–3444. https://doi.org/10.1109/TIFS.2021.3078327

[17] Cordero, C. G., Vasilomanolakis, E., Wainakh, A., Mühlhäuser, M., & Nadjm-Tehrani, S. (2021). On generating network traffic datasets with synthetic attacks for intrusion detection. ACM Transactions on Privacy and Security, 24(2), 8. https://doi.org/10.1145/3424155

[18] Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. 4th International Conference on Information Systems Security and Privacy (ICISSP), 108–116. https://doi.org/10.5220/0006639801080116

[19] Moustafa, N., & Slay, J. (2015). UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). 2015 Military Communications and Information Systems Conference (MilCIS), 1–6. https://doi.org/10.1109/MilCIS.2015.7348942

[20] Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. ACM Computing Surveys, 46(4), 44. https://doi.org/10.1145/2523813

[21] 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/10.1561/0400000042

[22] Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. 23rd ACM Conference on Computer and Communications Security (CCS), 308–318. https://doi.org/10.1145/2976749.2978318

[23] McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & Agüera y Arcas, B. (2017). Communication-efficient learning of deep networks from decentralized data. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), 1273–1282. https://doi.org/10.48550/arXiv.1602.05629

[24] Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., et al. (2021). Advances and open problems in federated learning. Foundations and Trends in Machine Learning, 14(1–2), 1–210. https://doi.org/10.1561/2200000083

[25] Islam, S. R., Eberle, W., Ghafoor, S. K., & Ahmed, M. (2022). A systematic literature review on security orchestration, automation, and response systems. Computers & Security, 122, 102910. https://doi.org/10.1016/j.cose.2022.102910

[26] Kinyua, J., & Awuah, L. (2021). AI/ML in security orchestration, automation and response: Future research directions. Intelligent Automation & Soft Computing, 28(2), 527–545. https://doi.org/10.32604/iasc.2021.016240

[27] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144. https://doi.org/10.1145/2939672.2939778

[28] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774. https://doi.org/10.48550/arXiv.1705.07874

[29] Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. Conference on Fairness, Accountability, and Transparency (FAT*), 220–229. https://doi.org/10.1145/3287560.3287596

[30] Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé, H., & Crawford, K. (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86–92. https://doi.org/10.1145/3458723

[31] Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

[32] Vielberth, M., Böhm, F., Fichtinger, I., & Pernul, G. (2020). Security operations center: A systematic study and open challenges. IEEE Access, 8, 227756–227779. https://doi.org/10.1109/ACCESS.2020.3045514

[33] Alahmadi, B. A., Axon, L., & Martinovic, I. (2022). 99% false positives: A qualitative study of SOC analysts' perspectives on security alarms. 31st USENIX Security Symposium, 2783–2800. https://doi.org/10.5555/3698900.3699051

[34] Apruzzese, G., Laskov, P., de Oca, E. M., Mallouli, W., Rapa, L. B., Grammatopoulos, A. V., & Di Franco, F. (2023). The role of machine learning in cybersecurity. Digital Threats: Research and Practice, 4(1), 8. https://doi.org/10.1145/3545574

[35] Kaissis, G. A., Makowski, M. R., Rückert, D., & Braren, R. F. (2020). Secure, privacy-preserving and federated machine learning in medical imaging. Nature Machine Intelligence, 2(6), 305–311. https://doi.org/10.1038/s42256-020-0186-1

[36] Politou, E., Alepis, E., Patsakis, C., Casino, F., & Alazab, M. (2020). Delegated content erasure in IPFS.

Future Generation Computer Systems, 112, 956–964. https://doi.org/10.1016/j.future.2020.06.037

[37] Rose, S., Borchert, O., Mitchell, S., & Connelly, S. (2020). Zero trust architecture (NIST SP 800-207). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-207

[38] Oliveira, N., Praca, I., Maia, E., & Sousa, O. (2021). Intelligent cyber attack detection and classification for network-based intrusion detection systems. Applied Sciences, 11(4), 1674. https://doi.org/10.3390/app11041674

[39] Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology, 10(2), 12. https://doi.org/10.1145/3298981

[40] Bonawitz, K., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., Ivanov, V., et al. (2019). Towards federated learning at scale: System design. Proceedings of Machine Learning and Systems, 1, 374–388. https://doi.org/10.48550/arXiv.1902.01046

[41] Dong, Y., Cordonnier, J. B., & Loukas, A. (2021). Attention is not all you need: Pure attention loses rank doubly exponentially with depth. Proceedings of the 38th International Conference on Machine Learning (ICML), 2793–2803. https://doi.org/10.48550/arXiv.2103.03404

[42] Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017). Membership inference attacks against machine learning models. IEEE Symposium on Security and Privacy, 3–18. https://doi.org/10.1109/SP.2017.41

[43] Hwang, R.-H., Peng, M.-C., Huang, C.-W., Lin, P.-C., & Nguyen, V. L. (2020). An unsupervised deep learning model for early network traffic anomaly detection. IEEE Access, 8, 30387–30399. https://doi.org/10.1109/ACCESS.2020.2973023

[44] Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. IEEE Signal Processing Magazine, 37(3), 50–60. https://doi.org/10.1109/MSP.2020.2975749

[45] Zhao, Y., Li, M., Lai, L., Suda, N., Civin, D., & Chandra, V. (2018). Federated learning with non-IID data. arXiv preprint. https://doi.org/10.48550/arXiv.1806.00582

[46] Wang, J., Liu, Q., Liang, H., Joshi, G., & Poor, H. V. (2020). Tackling the objective inconsistency problem in heterogeneous federated optimization. Advances in Neural Information Processing Systems, 33, 7611–7623. https://doi.org/10.48550/arXiv.2007.07481

[47] Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L.-M., Rothchild, D., So, D., Texier, M., & Dean, J. (2021). Carbon emissions and large neural network training. arXiv preprint. https://doi.org/10.48550/arXiv.2104.10350

[48] Madry, A., Makelov, A., Schmidt, L., Tsipras, D., & Vladu, A. (2018). Towards deep learning models resistant to adversarial attacks. International Conference on Learning Representations (ICLR). https://doi.org/10.48550/arXiv.1706.06083

[49] Wang, Z., Wang, N., Su, X., & Ge, S. (2020). An empirical study on business analytics affordances enhancing the management of cloud computing data security. International Journal of Information Management, 50, 387–394. https://doi.org/10.1016/j.ijinfomgt.2019.09.002

[50] Henson, B., Reyns, B. W., & Fisher, B. S. (2019). Cybercrime victimization. In D. Vazsonyi et al. (Eds.), The handbook of criminological theory (2nd ed.). Cambridge University Press.

https://doi.org/10.1017/9781108565684.016