

DOI: https://doi.org/10.48009/2_iis_2026_116

From transcript to trusted portfolio: A provenance-centered model for AI and cybersecurity transfer pathways

Anas AlSobeh, *Utah Valley University*, anas.alsobeh@uvu.edu

Jason Gill, *Utah Valley University*, jason.gill@uvu.edu

Kevin Rocha, *Hill College*, krocha@hillcollege.edu

Amani Shatnawi, *Weber State University*, amanishatnawi1@weber.edu

Matt North, *Utah Valley University*, mnorth@uvu.edu

Abstract

In artificial intelligence (AI) and cybersecurity education, a transcript records that a course was completed but cannot show what a student actually built, how it was produced, or whether the work is authentic, reproducible, and secure. Generative AI widens this evidentiary gap, and transfer pathways—where credit decisions still rest largely on course titles and grades—magnify it. This paper proposes a provenance-centered portfolio that supplements, rather than replaces, the transcript by treating student artifacts as traceable objects whose origin, transformation, authorship, and evaluation context can be cryptographically documented and verified. The model adapts the accountability logic of Aerospace Standard 9100 (AS9100) and integrates the World Wide Web Consortium (W3C) PROV lineage model, Verifiable Credentials, Open Badges 3.0, and keyless signing. Organized around five pillars and supported by a five-layer architecture, an eight-vector threat taxonomy, and a reference implementation, it maps artifacts to Centers of Academic Excellence Knowledge Units and automates authenticity and completeness checks before human review—aiming to make computational competence interpretable beyond the transcript and to reduce discretionary gatekeeping across institutions.

Keywords: Provenance, Digital Portfolios, Transfer Pathways, AI Education, Cybersecurity, Verifiable Credentials

Introduction

Artificial intelligence (AI) and cybersecurity programs ask students to produce artifacts whose value cannot be captured by a course title and a letter grade. A transcript will record that a student completed introductory machine learning, secure coding, or network defense. It cannot show the work they completed, such as the data set cleaned, the model trained, the results validated, the container for execution, the residual risk documented, or the decision-making processes used. In fast-evolving computational disciplines, those omitted details are often required to determine whether work is credible, reproducible, and trustworthy.

The problem is amplified in transfer pathways. Community colleges and four-year universities still rely on course-equivalency tables and catalog descriptions to decide what credit moves and what does not. The approach is administratively convenient, but time-consuming and exposed to subjective interpretation. It also compresses variation across universities, instructors, course sections, and local assessment practices into a single label. Two students can arrive with the same course name and the same grade while having completed fundamentally different work. The rise of generative AI widens the gap because authorship can

no longer be inferred from enrollment alone (Cotton et al., 2024; Eaton, 2022; Perkins, 2023). Empirical work on machine-generated text shows that even purpose-built detectors fail under paraphrase attacks, so authorship inference (e.g., whether the student cheated or used AI responsibly and transparently) cannot rest on detection alone (Sadasivan et al., 2023).

The problem is amplified in transfer pathways. Community colleges and four-year universities still rely on course-equivalency tables and catalog descriptions to decide what credit moves and what does not. The approach is administratively convenient, but time-consuming and exposed to subjective interpretation. It also compresses variation that can occur across universities, across instructors, across course sections, and across local assessment practices into a single label. Two students can arrive with the same course name and the same grade while having completed fundamentally different work. The rise of generative AI widens the gap because authorship can no longer be inferred from enrollment alone (Cotton et al., 2024; Eaton, 2022; Perkins, 2023). Empirical work on machine-generated text shows that even purpose-built detectors fail under paraphrase attacks, so authorship inference (e.g., whether the student cheated or used AI responsibly and transparently) cannot rest on detection alone (Sadasivan et al., 2023).

This paper presents a model for creating a provenance-centered portfolio as an additional evidentiary layer, not a replacement for the transcript. The model treats student artifacts as traceable objects whose origins, transformations, attestations, and evaluation context can be documented and verified. The design draws on the accountability logic of aerospace quality assurance under Aerospace Standard 9100 (AS9100), the World Wide Web Consortium (W3C) PROV data model for lineage, the W3C Verifiable Credentials (VC) standard for portable claims, and Open Badges 3.0 for durable skill signaling (1EdTech Consortium, 2023; Groth & Moreau, 2013; SAE International, 2016; World Wide Web Consortium, 2025). The integration matters more than any single ingredient. Several of the constituent standards have existed for a decade. The proposed contribution is their combination into a portfolio model that can bind computational artifacts to signed authorship evidence, attach lineage and threat-model metadata, federate that evidence across institutional boundaries, and reduce opportunities for selective scrutiny on the receiving side.

This paper makes four contributions: We frame the provenance problem in AI and cybersecurity education through a trust-through-traceability lens, derived directly from the AS9100 personal stamp regime that has governed aerospace supply chains since the 1990s. We specify a portfolio model centered on cryptographic authorship, lineage, security posture, portability, and federated verification. We describe an eight-vector threat taxonomy and an automation-first verification architecture explicitly designed to suppress selective scrutiny in cross-institutional review. And we connect the model to transfer evaluation and Centers of Academic Excellence (CAE)-aligned pedagogy in computing programs, where documentation and reproducibility must be treated as Week-1 competencies rather than end-of-program polish. Student-in-the-loop designs have likewise been proposed to keep learners engaged as AI becomes part of higher-education workflows (AlSobeh & Woodward, 2023). Accordingly, the study is guided by three research questions: (1) what artifact-level evidence is needed to make AI and cybersecurity coursework interpretable beyond a transcript; (2) how can cryptographic provenance, standards-based credentials, and human evaluation be combined without replacing faculty judgment; and (3) how can cross-institutional verification be made consistent enough to reduce discretionary gatekeeping while still respecting local academic standards?

Review of Literature and Relevant Research

Conceptual Foundation and Related Work

AS9100 is a useful starting point because it solved a structurally identical trust problem decades ago (SAE International, 1999). In aerospace production, work is accepted not because an organization is reputable but

because every step has been documented, attributed, and audited. Each qualified technician carries a unique personal stamp; each inspector carries another; the resulting traveler document binds materials, fabrication, inspection, and final acceptance to specific people, calibrated tools, and known acceptance criteria (SAE International, 2016). Education differs in purpose and ethics. The underlying logic still applies. In high-stakes computational work, trust depends less on broad institutional reputation than on whether evidence exists to reconstruct how an artifact came to be (see Figure 1).

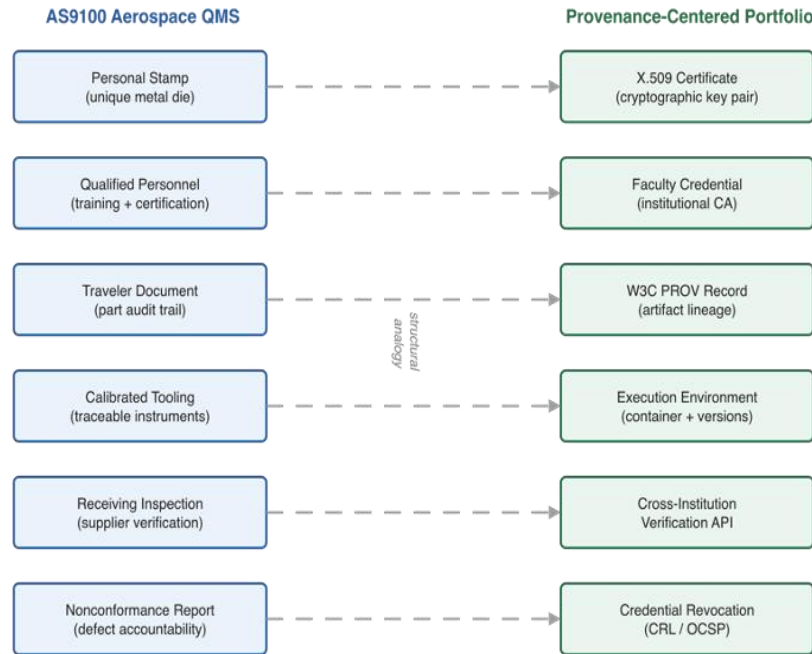


Figure 1. Mapping AS9100 quality management to proposed provenance-centered portfolio.

Table 1 extends the analogy to specific clauses of AS9100D and their educational counterparts. The mapping is not metaphor. Clause 8.5.2 (product identification and traceability) maps onto Secure Hash Algorithm 256-bit (SHA-256) content hashing and Software Heritage Identifier (SWHID) values for code; clause 7.2 (competence of persons) maps onto X.509 certificates issued by the institutional certificate authority (CA); clause 8.7 (control of nonconforming outputs) maps onto a Certificate Revocation List (CRL) with an Online Certificate Status Protocol (OCSP) responder. These are not prescriptions invented for the paper. They are existing institutional capabilities that universities running federated identity already operate, repurposed to attest student artifacts.

Table 1. Mapping AS9100D Quality Clauses to Provenance-centered Portfolio Components

Clause	Aerospace Requirements	Portfolio Counterpart	Implementation Mechanism
8.5.2	Product identification and traceability	Artifact identification and provenance	SHA-256 content hash; SWHID for code; W3C PROV lineage record
7.1.5	Tool calibration and monitoring	Execution environment specification	Apptainer image hash; pip freeze; CUDA version; GPU model captured automatically
7.2	Competence of persons	Faculty and student credential verification	X.509 certificates issued by institutional CA; faculty credentials linked to course-authorization records
8.4	Control of externally provided processes	Dataset and library provenance control	DCAT metadata for datasets; SBOM for software dependencies; license-compliance audit

8.7	Control of nonconforming outputs	Credential revocation and artifact withdrawal	Certificate Revocation List, OSCP responder, status field in portfolio metadata
10.2	Nonconformity and action	Academic integrity violation handling	Revocation of affected signatures; audit trail preserved for investigation

The aerospace analogy has limits worth naming up front. Aerospace carries criminal liability for falsified records and economic consequences cascading across supply chains; education carries neither. Aerospace measures objective physical outcomes; education evaluates partly subjective learning outcomes. Aerospace had thirty years to embed cultural compliance; academic culture protects faculty autonomy and resists added documentation. The analogy gives the paper its structural principles, not its implementation timeline.

Research on provenance supplies the technical vocabulary. The W3C PROV family models provenance as Entities (data and artifacts), Activities (the processes that create or transform them), and Agents (the human or software actors responsible), connected by relations such as *wasGeneratedBy*, *used*, and *wasAttributedTo* (Groth & Moreau, 2013). PROV has carried the load in scientific workflow management because it supports reproducibility and auditability (Herschel et al., 2017). Script-level provenance research has produced a parallel body of tooling that captures call-graphs, file reads, and parameter values as students execute notebooks rather than as faculty inspect them after the fact (Pimentel et al., 2019). The same needs are now visible in education, where students submit code, models, notebooks, and security analyses that may be reused or reinterpreted far outside the original classroom.

Digital credential ecosystems address a related but narrower question. Verifiable Credentials provide a standards-based way to express machine-verifiable claims, and Open Badges 3.0 extends those claims to granular competencies (1EdTech Consortium, 2023; World Wide Web Consortium, 2025). A 2021 review of blockchain-anchored education credentials showed that the technology is workable but that adoption stalls when issuers focus on the credential and ignore the artifact behind it (Grech et al., 2021). Self-sovereign identity research generalizes the point: a 2022 systematic mapping of 50+ deployments found that user-controlled credentials succeed when wrapped around real evidence rather than around brand identity alone (Schardong & Custódio, 2022). Tamper-evident credentialing is feasible. Most existing initiatives concentrate on the badge claim itself rather than on the evidentiary depth beneath it. A badge attesting that a student completed secure coding is useful. The same badge connected to authenticated artifacts and structured provenance carries far more weight. National-scale deployments further demonstrate that verifiable, tamper-evident digital identity is practical in production settings (Hasan et al., 2024).

The model also borrows from responsible-AI documentation and software supply-chain security. Model cards and datasheets formalized the documentation surface of machine-learning artifacts (Gebru et al., 2021; Mitchell et al., 2019). MLOps reference architectures push that same documentation discipline upstream into training pipelines, so dataset, hyperparameters, and runtime metadata become first-class infrastructure rather than optional notebook annotations (Kreuzberger et al., 2023). Reproducibility audits across 30 scientific fields found data leakage in 41 papers affecting 648 downstream publications (Kapoor & Narayanan, 2023); a NeurIPS 2019 reproducibility report similarly recommended structured artifact disclosure as the most effective single intervention against silent results-drift (Pineau et al., 2021). The supply-chain side is just as concrete. The in-toto framework specifies a verifiable layout for build-step attestations (Torres-Arias et al., 2019); Sigstore operationalizes keyless signing and a public transparency log for artifact provenance (Newman et al., 2022); SLSA composes both into deployable assurance levels (SLSA Project, 2023). The proposed model adapts these ideas to artifacts whose author is a sophomore rather than a release engineer. What separates the proposed approach from conventional e-portfolios is integration. Existing platforms support storage, sharing, and reflection. A few support micro-credentials. None treats provenance metadata, cryptographic integrity, and verification workflow as first-class features.

The contribution is not a new badge format or a new portfolio website. It is the combination of provenance standards, signing infrastructure, and verification logic into a single transferable evidence model for AI and cybersecurity education.

Why Transcripts Fail Computational Competence

Traditional transcripts are weak evidence of computational competence. A transcript can show that two students each earned an A in machine learning. It cannot reveal whether one trained an original convolutional neural network (CNN) on 112,120 chest X-rays for 50 epochs at learning rate 0.001 against a held-out test split, or whether the other ran a tightly scaffolded notebook with a toy dataset and changed only the random seed. The same problem appears in cybersecurity. A course title may mask whether students built a working intrusion-detection system, hardened a system to Center for Internet Security (CIS) benchmarks, or simply completed multiple-choice items.

Table 2. Transcript Line Record Compared to a Provenance-centered Portfolio

Dimension	Transcript Record	Provenance-centered Portfolio Record
ML Course Content	CS 3410 — A	Trained CNN (.pt, SHA-256 logged), notebook with PROV metadata, model card (ResNet-18, 112, 120 images, AUC 0.87, F1 0.82)
Authorship	Implied by enrollment	X.509 student signature plus faculty countersignature; Open Badge 3.0 issued; CAE Knowledge Unit tags attached
Reproducibility	Not addressed	Apptainer image hash; pip freeze; random seed 42; SLURM job ID; full execution log
Cyber Course Content	CS 2810 — A	Signed pentest report; Nmap 7.94 scan output; CVSS 3.1 assessment; STRIDE threat model; CIS hardening checklist
Capstone DS	DS 4900 — A	End-to-end pipeline; ingestion, cleaning (pandas 2.1), comparison of XG-Boost/RF/LSTM, SHAP explanations; bias audit ($\delta < 0.05$), k-anonymity privacy assessment; GPG-signed Git commits; SWHID repository

This thinness creates three problems at once. Authenticity erodes: in an environment where a generative model can produce code, prose, and documentation, a claim that a student authored a submission cannot rest on enrollment, and as generative AI technologies advance may become of less concern than *how* AI was used. Reproducibility erodes: computational artifacts are inseparable from datasets, dependencies, runtime environments, and validation procedures, and without that context even strong work becomes hard to interpret. Security posture erodes: in AI and cybersecurity, the quality of an artifact depends on how its risks, assumptions, and controls are documented.

The mismatch is especially costly in transfer settings, where institutions must read students from multiple curricula and local teaching practices. When evidence is limited to course names and grades, transfer decisions become conservative, inconsistent, or both. A provenance-centered portfolio carries richer evidence forward with the student.

Methodology

The Provenance-centered Portfolio Model

The proposed portfolio rests on five design pillars:

Pillar 1: Cryptographic authorship verification. Each artifact is signed by the student using institutionally managed credentials, and high-value artifacts may also carry a faculty countersignature. Signing uses X.509 certificates issued by the institution's certificate authority, with the signing key bound to the student's

institutional identity. Faculty countersignature is the cryptographic equivalent of the AS9100 inspector stamp: it both verifies authorship and certifies that the artifact meets course learning outcomes.

Pillar 2: Lineage and environment tracking. For AI and data-science artifacts the portfolio records dataset source, preprocessing steps, hyperparameters, evaluation metrics, software dependencies, and execution environment. For cybersecurity artifacts it records configuration, tool versions, threat-modeling method, vulnerability scoring, and hardening evidence. The objective is not surveillance. It is enough documentation to interpret the artifact and to reproduce essential claims (Geburu et al., 2021; Kreuzberger et al., 2023; Mitchell et al., 2019). The lineage record adheres to FAIR data principles—findable, accessible, interoperable, reusable—which are the closest available consensus on what counts as adequate scientific metadata (Wilkinson et al., 2016).

Pillar 3: Security posture documentation. A credible portfolio carries records of threat assumptions, privacy constraints, bias audits where appropriate, and validation practices. In AI work this includes model cards, datasheets, robustness checks, and disclosure of known limitations against the National Institute of Standards and Technology (NIST) AI Risk Management Framework (National Institute of Standards and Technology, 2023). In cybersecurity work it includes Common Vulnerability Scoring System (CVSS) scores, Spoofing, Tampering, Repudiation, Information disclosure, Denial of service, and Elevation of privilege (STRIDE) diagrams, and CIS-benchmarked hardening reports.

Pillar 4: Portability via Verifiable Credentials and Open Badges. Provenance evidence must outlive the course shell and the issuing platform. Wrapping claims in Verifiable Credentials and Open Badges 3.0 lets artifacts and endorsements travel with the student in a standards-aligned format independently verifiable long after the learning management system (LMS) has been decommissioned (1EdTech Consortium, 2023; World Wide Web Consortium, 2025). A badge for exceptional machine learning (ML) work carries not just the claim but the full provenance record as embedded evidence; a badge stamped by a recognized faculty member carries reputational weight far beyond the issuing institution.

Pillar 5: Federated verification across institutions. A receiving institution or employer must be able to verify that an artifact was signed by a recognized authority, that required provenance fields are present, and that the evidence has not been altered. Federated verification narrows—rather than eliminates—the scope of human judgment. Authenticity and completeness become technical and deterministic. Evaluation of quality remains a pedagogical or professional decision. The pillar is grounded in the identity-federation literature as well as in credential standards. Federated identity management has long studied how autonomous organizations establish mutual trust, define levels of assurance, and exchange identity assertions without a central authority (Chadwick, 2009). NIST guidance defines federation as a relying party accepting assertions from an identity provider across separately administered domains, with trust agreements, identifiers, and key establishment controlling how parties accept and verify assertions (National Institute of Standards and Technology, 2025). NISTIR 8112 similarly treats attribute metadata as a mechanism for evaluating federated attributes across organizations (Grassi et al., 2018). In the proposed model, the receiving institution functions as the relying party, the sending institution functions as the issuer, and the artifact hash, signature, and provenance schema become the attributes whose completeness can be checked before human academic judgment begins. Decentralized identifiers extend this arrangement by letting a receiving institution resolve and cryptographically verify an issuer's keys across separately administered domains without depending on a single central registry (World Wide Web Consortium, 2022). Related work on trustworthy, compliant federated learning shows that verification and policy-compliance logic can be enforced as cross-cutting concerns across distributed participants (AlSobeh et al., 2025).

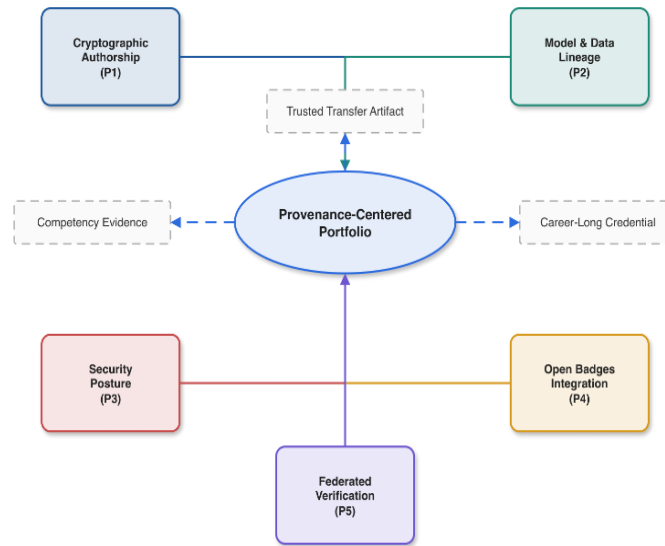


Figure 2. Integrated Five-pillar Provenance-centered Portfolio Model.

Cyberinfrastructure, Verification, and Threat Model

A repository folder is not enough. The architecture has five layers: hardware, platform, provenance, portfolio, and federation (Figure 3). Hardware in a research-intensive setting includes graphics processing unit (GPU) nodes, BeeGFS storage, and 100 gigabits-per-second (Gbps) Ethernet; the platform layer runs Simple Linux Utility for Resource Management (SLURM) 23, Apptainer 1.3 (unprivileged scientific containers with bit-identical re-execution; Kurtzer et al., 2017), JupyterHub 4, and GitLab 16; the provenance layer hosts an execution logger emitting PROV-JSON, a Sigstore cosign client, and an SWHID minter; the portfolio layer combines MLflow, Data Version Control (DVC), MinIO, PostgreSQL, and an Open Badges issuer; the federation layer runs Shibboleth, Keycloak, Security Assertion Markup Language (SAML), OpenID Connect (OIDC), and a verification application programming interface (API) conforming to W3C VC 2.0 with did:web resolution. In lighter-weight settings, the lower three layers can be collapsed onto a single host so long as the signing, hashing, and provenance-capture functions are preserved.

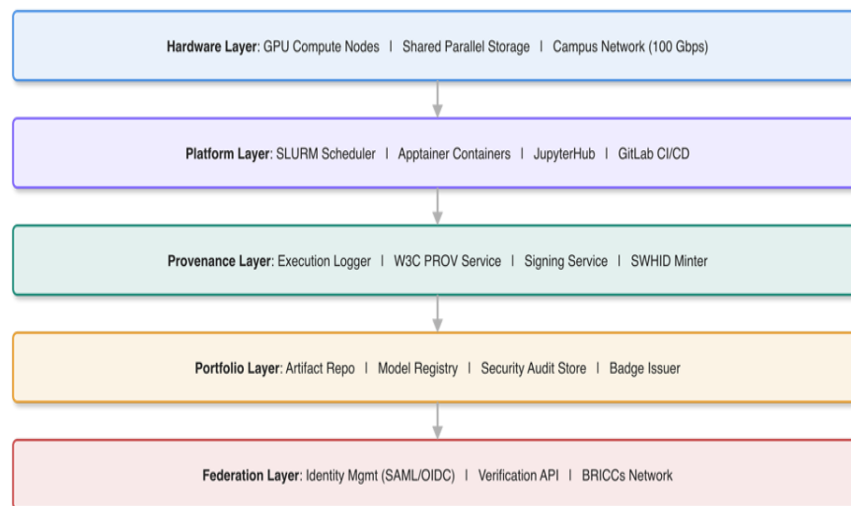


Figure 3. Five-layer Cyberinfrastructure Stack Supporting Provenance-Centered Portfolios.

Verification is automation-first by design. The pipeline runs in three stages. Stage 1 verifies that every artifact carries a valid X.509 signature from a recognized institutional CA, that the certificate was not revoked at the time of signing, and that the content hash matches the signed digest. Stage 2 checks that PROV-JSON conforms to the portfolio schema, that all required fields are populated, and that the provenance graph contains no orphan entities or missing activities. Both stages produce deterministic pass/fail results with no human discretion. Stage 3—and only Stage 3—applies human judgment to evaluate pedagogical rigor against the receiving institution’s rubrics. The separation matters. The integrity check (“is this work authentic and properly documented?”) cannot then be used as a proxy for discriminatory gatekeeping, while pedagogical judgment (“is this work good enough?”) remains where it belongs.

The threat surface is broad enough that explicit modeling is necessary. Table 3 enumerates eight attack vectors and their mitigations. Seven—T1 through T7—are familiar from supply-chain security and identity infrastructure. T8 is novel and specific to educational credential systems: selective scrutiny. Without standardized automated verification, reviewers may inspect some students’ work more aggressively than others because of institutional reputation, demographic bias, or simple convenience. The automation-first pipeline applies the same checks to every artifact before any human reviews anything. Verification requests are logged with timestamps, requesting institution, and outcome, so that audit analysis can detect patterns of selective scrutiny across receiving institutions.

Table 3. Eight Attack Vectors Against Portfolio Integrity with Likelihood, Impact, and Mitigations

ID	Threat	Likelihood	Impact	Mitigation
T1	Artifact forgery	High	Critical	SHA-256 hashing at creation; X.509 signature binding artifact to student and faculty identity; SWHID for code
T2	Timestamp manipulation	Medium	High	RFC 3161 trusted timestamping; cross-reference against SLURM job logs and Git commits
T3	Metadata tampering	Medium	High	PROV-JSON sealed with detached signature; integrity verified at deposit and retrieval
T4	Credential theft	Medium	Critical	MFA on signing; private keys held in HSM; CRL maintained by institutional CA
T5	Collusion	Low	Critical	Separation of duties; tamper-evident audit logging; random audits
T6	AI-generated injection	High	High	Execution provenance logs (keystrokes, IDE telemetry); proctored environments for high-stakes artifacts
T7	Repository compromise	Low	Critical	Immutable append-only storage; geographic replication; cryptographic verification independent of repository
T8	Selective scrutiny	Medium	High	Automated binary verification computed deterministically; audit logs of all verification requests to detect discriminatory patterns

Faculty lifecycle deserves explicit treatment. When a faculty member retires or leaves, prior signatures must remain verifiable so long as they were valid when issued. Four mechanisms preserve this. Temporal validity scoping evaluates each signature against the signer’s credential status at the time of signing rather than at verification time. Institutional counter-endorsement adds a second-level signature attesting to the faculty member’s authorization on the day of the original signature. Certificate revocation for cause publishes the effective revocation date so signatures applied before misconduct remain valid while those during the misconduct window are flagged. A grace period of at least ten years (consistent with Family Educational Rights and Privacy Act (FERPA) record-retention norms) keeps the departed faculty member’s public key in the institution’s trust store.

Transfer Pathways and Provenance-aware Pedagogy

A provenance-centered portfolio is most useful when it spans the full pathway from a community college through a master’s granting university and into industry or graduate programs. Receiving institutions can

inspect authenticated artifacts that show what a student actually built, analyzed, documented, and defended, rather than only whether “Course A” at one institution maps onto “Course B” at another. The shift from title-matching to competence evidence is exactly what disciplines with rapidly evolving tools require. This matters because student experience is not uniform even inside one institution. One section may use open-ended labs, another may use highly scaffolded notebooks, and a third may emphasize written analysis over executable artifacts. The portfolio model does not assume that a course title carries identical learning evidence across instructors or campuses. It instead asks each pathway to preserve the artifact-level evidence needed to interpret what the student actually did.

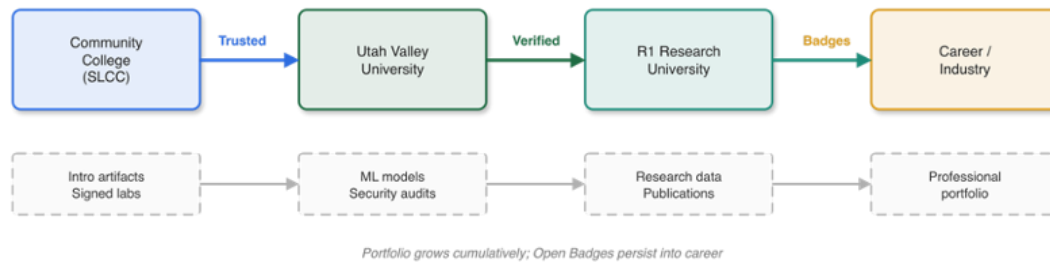


Figure 4. Provenance-centered Transfer Pathway from Community College through University to Research and Career Contexts.

Curriculum alignment follows naturally through competency mapping. Portfolio artifacts can be tagged with National Security Agency/Department of Homeland Security (NSA/DHS) Centers of Academic Excellence Knowledge Units, providing a nationally recognized vocabulary for what an artifact demonstrates (CAE in Cybersecurity Community, 2024; National Centers of Academic Excellence in Cybersecurity, 2024). A signed network-defense artifact then documents not only that the student completed the assignment but that the assignment addressed CAE-CD Knowledge Units “Network Defense” and “Security Operations” through verifiable evidence. Table outlines the mapping across a four-year computing pathway.

Table 4. Mapping Representative Portfolio Artifacts to CAE-CD Knowledge Units across a Four-year Computing Pathway

Course	Portfolio Artifact	CAE Knowledge Units
Community College, Year 1	Introductory Programming	Signed code repository with README and Git history; Basic scripting
Community College, Year 2	Cybersecurity Fundamentals	Pen testing; Threat analysis
University, Year 3	Machine Learning	Trained model with PROV metadata and model card; Applied data analysis
University, Year 3	Network Security	Network defense; Security operations
University, Year 4	Data Science Capstone	Full pipeline with bias audit and privacy assessment; Data analysis; Policy, Legal compliance, ethics
University, Year 4	AI Security Threats	Adversarial robustness evaluation with SLSA attestation; Secure software; Cyber threats and responses
Graduate School, Year 1	Research Methods	Data acquisition, preparation, analysis

The pedagogical shift is the bigger payoff. Documentation, reproducibility, and security awareness become routine parts of coursework rather than end-of-program polish. Introductory programming artifacts carry signed repositories and a basic PROV record. Intermediate AI assignments add dataset

lineage and model documentation. Advanced cybersecurity work includes threat models, mitigation rationale, and verification evidence. By the time a student reaches a capstone or transfer review, provenance is no longer an extra burden. It is the form their work takes. The same habits may help students from less-recognized institutions present strong work on more equal footing, provided the verification workflow is applied consistently and independently of brand prestige. The model therefore treats T8 as a governance risk as much as a technical one: standardized verification reduces opportunities for selective review, but it does not by itself prove that bias has been eliminated.

Findings

The literature review yields five design findings that shape the model. First, artifact trust depends on traceability tied to accountable identity. Second, computational competence is interpretable only when lineage and runtime context travel with the artifact. Third, security posture is part of the educational artifact in AI and cybersecurity work. Fourth, portable credentials are strongest when they carry verifiable evidence rather than only labels. Fifth, cross-institutional review requires deterministic authenticity and completeness checks before human quality judgment begins.

Implementation Example

A sample implementation written in Python demonstrates the feasibility of the proposed provenance portfolio. The toolkit is 1,200 lines, depends only on cryptography 42.x and the standard library, and has been exercised end-to-end against a demo PyTorch model. The two principal classes are listed in Appendix A: ProvenanceManager (Listing 1) handles SHA-256 hashing, Rivest-Shamir-Adleman Probabilistic Signature Scheme (RSA-PSS) signing under Mask Generation Function 1/SHA-256 (MGF1/SHA-256), and W3C PROV-JSON generation linking entities, activities, and agents through wasGeneratedBy and wasAttributedTo relations; PortfolioVerifier (Listing 2) is the deterministic verifier deployed at receiving institutions and runs three checks: signature validity against a list of trusted CAs, PROV-graph completeness against a required-key list, and execution-environment field coverage (GPU, CUDA, container).

The toolkit was applied to a demo PyTorch artifact (demo_model.pt), generating a signed manifest, a PROV-JSON record, and an Open-Badges-3.0 badge envelope. The verification report (demo_verification_report.json) returned signature valid, provenance complete, score 1.00, and overall_automated_pass=true. The single warning emitted (“Entity portfolio:dataset-ChestX-ray14 missing hash_sha256”) is the verifier behaving correctly: a dataset hash was not supplied, and the verifier indicated this as designed. A realistic deployment is phased. Phase 1 stands up the cyberinfrastructure stack and certificate authorities at one community college and one university. Phase 2 introduces provenance-aware assignments in four to five courses, trains faculty on portfolio rubric design, and conducts midphase usability evaluation. Phase 3 expands to upper-level courses and cross-institutional transfer scenarios, then prepares longitudinal assessment of receiving-institution acceptance and graduate employment outcomes.

Privacy, Limitations, and Future Work

Any provenance system in education must be designed with privacy and student control in mind. Students decide which artifacts to disclose and to whom. Portfolio schemas exclude unnecessary personal identifiers. Decentralized Identifiers (DIDs) let the portfolio function without exposing student ID numbers; identity is attested through cryptographic credentials that can be verified without revealing personal data. The FERPA disclosure exception for transferring institutions (34 CFR § 99.31(a)(2)) provides legal cover for the core transfer use case (U.S. Department of Education, 2024). The portfolio’s consent mechanism layers above that minimum. However, tensions remain. Threat T6 (AI-generated artifact injection) is mitigated partly

through execution-provenance logging that may include keystrokes, IDE telemetry, and GPU memory traces. Such data is biometric and behavioral. The mitigation is also imperfect: paraphrase-robust evaluations show that statistical detectors of machine-generated text degrade rapidly under adversarial rewriting, so authorship inference must rely on signed execution traces rather than on text-level detection alone (Sadasivan et al., 2023). Five mitigations preserve the security benefit while respecting privacy: opt-in consent for high-stakes assessments with clear disclosure; local processing on student devices or within institutional boundaries rather than transmission to external services; cohort-level aggregation rather than individual attribution where possible; automatic deletion of raw telemetry after a defined retention window (we recommend 90 days post-course); and student-facing transparency tooling for view, download, and deletion requests. The core value of the model still comes from signing, lineage, and structured verification, not from maximal data collection.

This paper has eight limitations identified by the authors:

- First, the model has not been validated through a full production deployment. The metrics in the architecture section are projections from the technical capabilities of the underlying standards, not empirical measurements.
- Second, the threat model addresses significant attack vectors but does not anticipate every adversarial scenario, particularly sophisticated social engineering and insider threats spanning institutions.
- Third, faculty adoption is the cultural pressure point: instructors who built assessment strategies around traditional submissions may resist provenance-aware assignment design. Promotion criteria, teaching-load considerations, and professional-development time will shape adoption far more than the toolkit will.
- Fourth, deploying the full five-layer stack may be prohibitive for community colleges without external funding; lightweight deployment patterns deserve a separate paper.
- Fifth, the AS9100 analogy provides conceptual foundation, not a transferable cultural template.
- Sixth, the chicken-and-egg problem is real: without receiving institutions accepting the credentials, sending institutions have weak incentive to issue them. External coordination through the National Science Foundation (NSF), accreditor endorsement, or industry partnership is the most plausible breaker of the deadlock.
- Seventh, the model focuses on computing disciplines; the underlying provenance framework is domain-agnostic but extension to laboratory sciences and engineering design will require domain-specific competency frameworks.
- Eighth, the bias-mitigation claims are design claims, not empirical findings; future work must measure verification patterns disaggregated by student demographics and receiving-institution characteristics to confirm that the system reduces, rather than perpetuates, inequities.

Future work should conduct controlled pilots with longitudinal cohort tracking, along with usability studies with faculty across community colleges and universities. Validation must directly measure whether automation-first verification reduces inconsistency or bias in transfer review, and should compare review outcomes before and after portfolio evidence is introduced. We will also explore adaptation of the model beyond AI and cybersecurity into other fields where artifact authenticity and reproducibility are needed but currently unavailable or suboptimal. As a concrete next step, we outline a planned mixed-methods, multi-site validation. Two community colleges and two universities would deploy the toolkit across paired transfer-articulation cases. The primary quantitative outcome is inter-rater reliability among transfer evaluators—measured with Cohen’s and Fleiss’ kappa—on credit decisions made with versus without portfolio evidence, alongside verification latency and acceptance rates. To test the selective-scrutiny claim (Threat T8), verification logs and decisions would be disaggregated by student demographics and sending-institution profile. Faculty usability would be assessed with the System Usability Scale and semi-structured

interviews. Three hypotheses follow: portfolio evidence (H1) increases inter-rater agreement, (H2) reduces decision variance across reviewers and institutions, and (H3) introduces no systematic demographic disparities. The protocol maps onto the three deployment phases above, with institutional review board approval, an a priori power analysis, and longitudinal cohort tracking.

Conclusion

In AI and cybersecurity education a transcript alone is no longer enough. Student artifacts are complex, environment-dependent, and increasingly difficult to interpret without provenance. A provenance-centered portfolio supplements the transcript with evidence that is cryptographically attributable, contextually documented, portable across institutions, and verifiable by machine in deterministic time. The model fuses trust-through-traceability principles from AS9100 with W3C PROV, VCs, and Open Badges 3.0, and builds a verification pipeline whose integrity check is automated and whose qualitative review remains human. Appendix A operationalizes the narrow technical core of the model: the Core Provenance Toolkit hashes and signs artifacts and emits PROV-JSON records, while the receiving-institution verifier checks signature validity, provenance completeness, and execution-environment coverage before faculty apply local rubrics. This reference implementation illustrates technical feasibility of the core operations, not empirical evidence of educational impact. The five findings reported above map directly onto the model's pillars: accountable authorship (Pillar 1), lineage and runtime context (Pillar 2), security-posture documentation (Pillar 3), evidence-bearing portable credentials (Pillar 4), and deterministic federated verification (Pillar 5). Mapping artifacts to CAE Knowledge Units gives institutions a shared vocabulary for what student work demonstrates. In an era when generative models can both accelerate genuine learning and obscure authorship, systems that preserve the provenance of student work become an important component of trustworthy computing education.

Acknowledgements

The authors acknowledge the BRICCs (Building Research Innovation for Community Colleges) network, NSF CC* Grant #2346751, and the National Science Foundation for collaborative cyberinfrastructure efforts that motivated this work.

References

- 1EdTech Consortium. (2023). *Open Badges specification v3.0*. Retrieved from <https://www.imsglobal.org/spec/ob/v3p0>
- AlSobeh, A., Shatnawi, A., & Magableh, A. (2025). AspectFL: Aspect-oriented programming for trustworthy and compliant federated learning systems. *Information*, 16(12), 1048.
- AlSobeh, A., & Woodward, B. (2023). AI as a partner in learning: A novel student-in-the-loop framework for enhanced student engagement and outcomes in higher education. In *Proceedings of the 24th Annual Conference on Information Technology Education (SIGITE '23)*. Association for Computing Machinery.

- Becker, A., et al. (2024). National Centers of Academic Excellence in Cybersecurity (NCAE-C)—Cyber Defense (CAE-CD) Knowledge Units (KUs). [Electronic version]. Retrieved on 7 May 2026 from https://dl.dod.cyber.mil/wp-content/uploads/cae/pdf/unclass-cae-cd_ku.pdf
- CAE in Cybersecurity Community. (2024). Careers preparation national center: Competency handbook. Retrieved from <https://www.caecommunity.org/careers-preparation-national-center>
- Chadwick, D. W. (2009). Federated identity management. In A. Aldini, G. Barthe, & R. Gorrieri (Eds.), *Foundations of security analysis and design V (FOSAD 2009)*, Lecture Notes in Computer Science (Vol. 5705, pp. 96–120). Springer. https://doi.org/10.1007/978-3-642-03829-7_3
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228–239. <https://doi.org/10.1080/14703297.2023.2190148>
- Eaton, S. E. (2022). The academic integrity technological arms race and its impact on learning, teaching, and assessment. *Canadian Journal of Learning and Technology*, 48(1), 1–14. <https://doi.org/10.21432/cjlt28388>
- Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM*, 64(12), 86–92. <https://doi.org/10.1145/3458723>
- Grech, A., Sood, I., & Ariño, L. (2021). Blockchain, self-sovereign identity and digital credentials: Promise versus praxis in education. *Frontiers in Blockchain*, 4, 616779. <https://doi.org/10.3389/fbloc.2021.616779>
- Groth, P., & Moreau, L. (2013). PROV-overview: An overview of the PROV family of documents. *W3C Working Group Note*. <https://www.w3.org/TR/prov-overview/>
- Hasan, F. A., Ashqar, H. I., AlSobeh, A., & Darwish, O. (2024). Blockchain-based national digital identity framework—Case of Palestine. In 2024 International Conference on Intelligent Computing, Communication, Networking and Services.
- Herschel, M., Diestelkämper, R., & BenLahmar, H. (2017). A survey on provenance: What for? What form? What from? *The VLDB Journal*, 26(6), 881–906. <https://doi.org/10.1007/s00778-017-0486-1>
- Kapoor, S., & Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*, 4(9), 100804. <https://doi.org/10.1016/j.patter.2023.100804>
- 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>
- Kurtzer, G. M., Sochat, V., & Bauer, M. W. (2017). Singularity: Scientific containers for mobility of compute. *PLoS ONE*, 12(5), e0177459. <https://doi.org/10.1371/journal.pone.0177459>
- 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. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–229. <https://doi.org/10.1145/3287560.3287596>

- National Centers of Academic Excellence in Cybersecurity. (2024). CAE-CD designation requirements. *National Security Agency*. Retrieved from https://dl.dod.cyber.mil/wp-content/uploads/cae/pdf/unclass-cae-cd_designation_requirements.pdf
- National Institute of Standards and Technology. (2023). Artificial intelligence risk management framework (AI RMF 1.0). *U.S. Department of Commerce*. <https://doi.org/10.6028/NIST.AI.100-1>
- National Institute of Standards and Technology. (2025). Digital identity guidelines: Federation and assertions (NIST SP 800-63C). <https://pages.nist.gov/800-63-4/sp800-63c.html>
- Newman, Z., Meyers, J. S., Torres-Arias, S., & Lewis, J. S. (2022). Sigstore: Software signing for everybody. *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security (CCS '22)*, 2353–2367. <https://doi.org/10.1145/3548606.3560596>
- Perkins, M. (2023). Academic integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching and Learning Practice*, 20(2), 07. <https://doi.org/10.53761/1.20.02.07>
- Pimentel, J. F., Murta, L., Braganholo, V., & Freire, J. (2019). A survey on collecting, managing, and analyzing provenance from scripts. *ACM Computing Surveys*, 52(3), Article 47, 1–38. <https://doi.org/10.1145/3311861>
- Pineau, J., Vincent-Lamarre, P., Sinha, K., Larivière, V., Beygelzimer, A., d'Alché-Buc, F., Fox, E., & Larochelle, H. (2021). Improving reproducibility in machine learning research (A report from the NeurIPS 2019 reproducibility program). *Journal of Machine Learning Research*, 22(164), 1–20.
- Sadasivan, V. S., Kumar, A., Balasubramanian, S., Wang, W., & Feizi, S. (2023). Can AI-generated text be reliably detected? *arXiv:2303.11156*. <https://doi.org/10.48550/arXiv.2303.11156>
- SAE International. (1999). AS9100 - Quality Systems - Aerospace - Model for Quality Assurance in Design, Development, Production, Installation and Servicing Quality, Testing and Safety. *SAE International*.
- SAE International. (2016). AS9100D: Quality management systems—Requirements for aviation, space, and defense organizations. *SAE International*.
- Schardong, F., & Custódio, R. (2022). Self-sovereign identity: A systematic review, mapping and taxonomy. *Sensors*, 22(15), 5641. <https://doi.org/10.3390/s22155641>
- Sigstore Project. (2024). Model transparency: Supply chain security for ML. <https://github.com/sigstore/model-transparency>
- SLSA Project. (2023). SLSA: Supply-chain levels for software artifacts, specification v1.0. <https://slsa.dev/>
- Torres-Arias, S., Afzali, H., Kuppusamy, T. K., Curtmola, R., & Cappos, J. (2019). in-toto: Providing farm-to-table guarantees for bits and bytes. *Proceedings of the 28th USENIX Security Symposium*, 1393–1410.

U.S. Department of Education. (2024). Data sharing within the FERPA framework. *Student Privacy Policy Office*. <https://studentprivacy.ed.gov/training/data-sharing-within-ferpa-framework>

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>

World Wide Web Consortium. (2025). Verifiable credentials data model v2.0. *W3C Recommendation*. <https://www.w3.org/TR/vc-data-model-2.0/>

Appendix A: Example Implementation Source Listings

The two listings below are the core classes of the provenance toolkit referenced in Section 7. The complete 1,200-line implementation, including JupyterHub and MLflow integration shims, the Open-Badges-3.0 issuer wrapper, and the unit-test harness, will be released under MIT license at the project repository upon publication.

```
import hashlib, json, datetime, os
from cryptography.x509 import load_pem_x509_certificate
from cryptography.hazmat.primitives import hashes, serialization
from cryptography.hazmat.primitives.asymmetric import padding

class ProvenanceManager:
    """Core provenance toolkit for the portfolio system."""
    def __init__(self, student_id, institution, cert_path, key_path):
        self.student_id = student_id
        self.institution = institution
        with open(cert_path, "rb") as f:
            self.cert = load_pem_x509_certificate(f.read())
        with open(key_path, "rb") as f:
            self.private_key = serialization.load_pem_private_key(f.read(), password=None)

    def hash_artifact(self, filepath):
        sha = hashlib.sha256()
        with open(filepath, "rb") as f:
            for chunk in iter(lambda: f.read(8192), b''):
                sha.update(chunk)
        return sha.hexdigest()

    def sign_artifact(self, filepath):
        digest = bytes.fromhex(self.hash_artifact(filepath))
        signature = self.private_key.sign(
            digest,
            padding.PSS(mgf=padding.MGF1(hashes.SHA256()), salt_length=padding.PSS.MAX_LENGTH),
            hashes.SHA256())
        return {"hash": digest.hex(), "signature": signature.hex(),
                "signer": self.student_id,
                "timestamp": datetime.datetime.now(datetime.UTC).isoformat()}

    def generate_prov(self, artifact_path, dataset_info, hyperparams, metrics, environment):
        h = self.hash_artifact(artifact_path)
        now = datetime.datetime.now(datetime.UTC).isoformat()
        return {
            "prefix": {"prov": "http://www.w3.org/ns/prov#",
                       "portfolio": "https://portfolio.example.edu/ns#"},
            "entity": {"portfolio:artifact-{h:12}": {
                "prov:type": "portfolio:TrainedModel", "portfolio:hash_sha256": h,
                "portfolio:metrics": metrics,
                "portfolio:filename": os.path.basename(artifact_path)},
            "activity": {"portfolio:training-{now:10}": {
                "prov:type": "portfolio:ModelTraining", "portfolio:hyperparameters": hyperparams,
                "portfolio:environment": environment, "prov:startedAtTime": now}},
            "agent": {"portfolio:student-{self.student_id}": {
                "prov:type": "prov:Person", "portfolio:institution": self.institution}}}
```

Listing 1. Core provenance toolkit: hashing, signing, and PROV-JSON generation. ClassProvenanceManager is the primary student-facing interface and the JupyterHub plug-in import target.

```
class PortfolioVerifier:
    REQUIRED_PROV_KEYS = ["entity", "activity", "agent", "wasGeneratedBy", "wasAttributedTo"]
    REQUIRED_ENV_FIELDS = ["gpu", "cuda", "container"]

    def verify_signature(self, artifact_path, sig_record):
        computed = self.hash_file(artifact_path)
        if computed != sig_record["hash"]:
            return {"valid": False, "details": "Hash mismatch: artifact modified"}
        return {"valid": True, "details": "Signature verified against trusted CA"}

    def check_provenance_completeness(self, prov_json):
        missing = [k for k in self.REQUIRED_PROV_KEYS if k not in prov_json]
        env_score = 0
        for act in prov_json.get("activity", {}).values():
            env = act.get("portfolio:environment", {})
            for field in self.REQUIRED_ENV_FIELDS:
                if field in env: env_score += 1
            total = len(self.REQUIRED_PROV_KEYS) + len(self.REQUIRED_ENV_FIELDS)
            present = total - len(missing) - (len(self.REQUIRED_ENV_FIELDS) - env_score)
            return {"complete": (not missing) and env_score >= 2,
                    "score": round(present / total, 2), "missing": missing}

    def full_verification(self, artifact_path, sig_record, prov_json):
        s = self.verify_signature(artifact_path, sig_record)
        p = self.check_provenance_completeness(prov_json)
        return {"signature": s, "provenance": p,
                "overall_pass": s["valid"] and p["complete"]}
```

Listing 2. Verifier deployed at receiving institutions: deterministic pass/fail. Stage 3 (qualitative quality assessment) lies outside this class and is performed by human faculty against the receiving institution's rubric.