

Preserving Learning Signals in AI-Supported Cybersecurity Education

Christopher Nott
cnott@inpraxia.ai
Independent Researcher
Tennessee, USA

Abstract

Cybersecurity instructors often judge more than a final answer. In an incident report, threat brief, policy memo, or lab reflection, the useful evidence may be the artifact a student selected, the hypothesis the student rejected, the uncertainty the student named, or the reason one risk received priority over another. Generative artificial intelligence (AI) can weaken that evidence by turning partial understanding into polished prose. Detection asks where prose came from; assessment asks what the student can do. This paper calls that problem learning signal compression, defined as the loss or thinning of evidence that instructors use to infer student competence. The concept extends assessment validity concerns into the specific setting of AI-supported cybersecurity assignments. Validity asks whether the evidence supports the intended interpretation of student performance. Learning signal compression identifies one way that evidence can become less trustworthy when AI hides the student's reasoning path behind a fluent artifact. The paper develops a signal-preservation review model for adult and postsecondary cybersecurity education. The model asks faculty and program leaders to examine four conditions before AI is permitted, required, or embedded in a course task. The conditions are instructional purpose, faculty judgment, assessment integrity, and governance support. Here assessment integrity refers to the strength of the inference from evidence to competence, not only to student honesty. The discussion applies the model to threat analysis, incident response reporting, and program-level AI policy. The contribution is a course-design aid for revising cybersecurity assignments so that AI-supported work remains assessable.

Keywords: cybersecurity education, generative AI, assessment validity, learning evidence, academic integrity, faculty judgment

Preserving Learning Signals in AI-Supported Cybersecurity Education

Christopher Nott

1. INTRODUCTION

Cybersecurity instructors have always had to read past the surface of student work. A report can use the right terms and still miss the investigation, a lab reflection can sound confident while skipping where the student got stuck, and a policy memo can cite a framework without showing whether the student understood the risk. These problems are familiar. Generative artificial intelligence (AI) makes them harder to see.

The issue shows up quickly in applied assignments. A student working through a ransomware exercise may collect several screenshots, copy a few log entries, and ask an AI system to produce an incident report. The result can be readable and technically plausible. It may identify an attack path, recommend reasonable controls, and use the professional tone expected in the assignment. What it may not show is how the student moved from evidence to conclusion.

That missing movement is central to cybersecurity education. Students are expected to interpret incomplete evidence, notice weak signals, reject poor explanations, prioritize risk, and defend recommendations. Those abilities are closer to professional judgment than to recall. They are also difficult to assess from a finished document alone, especially when AI can improve the document without improving the analysis behind it.

Learning signal compression names the loss when AI-supported work hides or thins the evidence an instructor needs to judge competence. In a cybersecurity course, a learning signal may be a selected artifact, a rejected hypothesis, a confidence rating, a timeline revision, or an answer to a follow-up question. When AI reduces that trail to a polished final product, the assignment may look stronger while producing weaker evidence.

This concept differs from plagiarism and from general assessment validity. Plagiarism concerns the improper presentation of another source's work as one's own. Learning signal compression can occur even when AI use is permitted and disclosed.

Current institutional responses often begin somewhere else. Policies ask whether students may use AI, whether use must be disclosed, whether AI-generated text must be cited, and whether instructors may rely on detection tools. These have administrative value but do little to help an instructor decide whether an assignment still works. A disclosed AI-supported report can remain a weak assessment if it leaves no visible trail of evidence. A prohibited AI use can be difficult to prove and may say less about the student than about the task design.

Cybersecurity programs should examine the evidence base of AI-supported assignments before considering tool novelty. Faculty need to decide which part of the student's reasoning must remain visible. The review model developed here turns that decision into four questions: whether the task names a cybersecurity judgment, whether faculty can evaluate the evidence trail, whether the assessment preserves enough evidence for a fair judgment, and whether governance supports the arrangement. The paper therefore contributes a conceptual vocabulary and a review model rather than empirical validation.

The question guiding this paper is how cybersecurity programs can preserve assessable evidence of student learning when AI tools become part of instructional and assessment work. The answer offered here begins with the evidence an assignment leaves behind.

2. THE ASSESSMENT PROBLEM IN CYBERSECURITY COURSES

Many cybersecurity courses rely on professional artifacts. Students write incident reports, intelligence briefs, vulnerability findings, risk memos, policy analyses, and post-lab reflections. Those artifacts support strong pedagogy when tied to authentic judgment. The difficulty is that the artifact is often cleaner than the cognition it represents. A report can hide how the student moved from packet capture to claim, from log entry to timeline, or from vulnerability description to remediation priority.

Assessment theory has long treated validity as an evidentiary problem. Instructors interpret performance and use that interpretation to make claims about learning. Evidence-centered design makes this chain explicit by connecting the claim about competence, the evidence needed to support the claim, and the task likely to elicit that evidence (Mislevy et al., 2003). Kaldaras et al. (2024) apply related concerns to generative AI assessment, arguing that it requires careful attention to the evidence supporting score interpretation. Their point applies to cybersecurity education. When the work product no longer shows the response process, the instructor has weaker grounds for judging the underlying construct.

The AI Assessment Scale (AIAS) is a useful starting point, asking instructors to define acceptable AI use in relation to learning outcomes (Perkins et al., 2024). That approach moves beyond the shallow choice between prohibition and permission. Cybersecurity education needs an additional layer. It must ask which parts of the student's reasoning must remain inspectable. A course may allow AI for brainstorming, critique, or report formatting while still requiring a separate evidence log, decision journal, or oral defense because those artifacts carry signals that the final report cannot provide.

Academic integrity belongs in this discussion, although it cannot carry the full assessment problem. AI text detection is a weak foundation for course design. Research on GPT detectors has documented misclassification problems, including risks for non-native English writers (Liang et al., 2023). Even when a detector performs well in a narrow testing condition, it does not tell the instructor whether a student can explain the event sequence, justify a severity rating, or recognize a false equivalence in a framework mapping. Detection asks where prose came from. Assessment asks what the student can do.

A more durable response redesigns the evidence environment around the task. A polished final artifact is one layer of evidence. The instructor also needs artifacts that make the analytic path visible, such as extracted indicators, rejected hypotheses, tool output, and timestamped drafts. These artifacts do not remove misconduct, but make learning harder to fake and easier to judge fairly.

3. LEARNING SIGNAL COMPRESSION

Learning signal compression is narrower than the broad claim that AI complicates assessment. It refers to a particular loss of traceable evidence. A student may use AI with permission and still produce work that tells the instructor too little. Cybersecurity assignments are vulnerable because they reward professional presentation. A fluent incident report, executive summary, or policy memo may be fine as a workplace artifact, but as assessment evidence each can be incomplete without the reasoning trail that gives it meaning. This loss is a form of cognitive offloading, in which reasoning is delegated to a tool and its record does not survive in the finished product (Risko & Gilbert, 2016).

Several forms of compression appear in AI-supported cybersecurity tasks. Surface fluency compression has AI improve organization and tone enough to mask weak analysis. Source compression occurs when the student's own evidence selection is replaced by a generated summary that gives no clear account of what was noticed, ignored, or weighted. Uncertainty compression turns tentative evidence into confident language. Framework compression invokes MITRE ATT&CK, the NIST Cybersecurity Framework (CSF), the National Initiative for Cybersecurity Education (NICE) Framework, or a compliance standard so it looks authoritative without showing why the mapping is justified (MITRE, n.d.; National Institute of Standards and Technology, 2024; Petersen et al., 2020). Sequence compression presents investigation steps as a neat timeline after the fact, with no record of the false starts that shaped the conclusion. These forms are analytical categories that emerged from the conceptual synthesis, not empirically validated types. The list is illustrative, not exhaustive, and the model that follows is organized around evidence conditions instead of the individual forms.

Consider a port-scanning exercise. A student who submits only a final paragraph saying the traffic indicates reconnaissance has given the instructor a conclusion. The stronger evidence would show which packets were examined, what ordinary traffic was ruled out, why the pattern suggested scanning, which alternative explanations were considered, and how confident the student is in the claim. AI may improve the paragraph. It cannot replace the evidence needed to evaluate the student's analytic work.

A useful assessment design begins with an inventory of the signal that matters most for the task. If the task is threat analysis, the signal may be evidence selection and hypothesis ranking. If the task is incident response, the signal may be sequencing and tradeoff reasoning. If the task is policy analysis, the signal may be interpretation under local constraints. If the task is vulnerability management, the signal may be severity reasoning, with the prose of the finding treated as secondary. Once the signal is named, AI use can be shaped around it instead of handled through blanket permission or prohibition.

4. DESIGN BASIS AND SCOPE

The model presented here is a design synthesis. It has not been tested as a classroom experiment. It brings together assessment validity, evidence-centered design, authentic assessment, cybersecurity curriculum guidance, responsible AI governance, and recent work on generative AI assessment. The synthesis is needed because each body of work addresses part of the problem while leaving assignment-level evidence decisions underdeveloped. These bodies of work were chosen for their direct bearing on assessment evidence and on AI use in cybersecurity coursework, and the account that follows is interpretive rather than systematic.

Assessment validity provides the main anchor. As Section 2 noted, evidence-centered design links claim, evidence, and task (Mislevy et al., 2003). That logic is essential, but generative AI changes the response process in ways that standard assignment designs may not anticipate. A completed report can appear to satisfy the task while concealing the route from evidence to claim. Learning signal compression gives a name to that route-level loss.

The distinction from existing validity concepts is important for the contribution of this paper. Validity is the larger interpretive standard. It asks whether a score or judgment is warranted. Within that standard, validity theory already treats the response processes behind a performance as a source of evidence (Messick, 1995). Learning signal compression describes one mechanism that can weaken the warrant in AI-supported cybersecurity tasks. The concern reaches beyond the student receiving help. The deeper issue is that the help may remove evidence of the student's own evidence selection, uncertainty management, and reasoning under pressure. A validity argument can identify the problem in theory. The review model proposed here gives faculty a practical way to locate the problem in an assignment.

As noted above, the AI Assessment Scale defines permitted AI use in relation to learning outcomes (Perkins et al., 2024), but does not fully specify which cybersecurity learning signals must remain visible when AI is allowed. A level of permitted AI use may be appropriate for one part of a task and inappropriate for another. A threat analysis assignment may allow AI critique after evidence selection, but restrict AI during the initial interpretation of logs. The model in this paper addresses that finer design decision.

Authentic assessment research adds another layer. Authentic tasks put students in realistic situations, constraints, and judgment demands. Recent scholarship argues that authenticity in assessment should include the practices and expectations that make performance meaningful in context (Ajjawi et al., 2024). Cybersecurity training research makes the same point through cyber ranges and realistic exercises, where learners must act under conditions that resemble operational pressure (Glas et al., 2023). Authentic-looking artifacts can still be weak evidence if AI produces the professional surface while the student's analytic process stays hidden.

Curriculum frameworks explain why the evidence question matters. Cybersecurity Curricula 2017 (CSEC2017) presents the field as connecting systems, software, human behavior, organizational risk, and policy (Joint Task Force on Cybersecurity Education, 2017), and the NICE Framework describes the work through tasks, knowledge, skills, and work roles (Petersen et al., 2020). Both place competence

in situated performance, so a cybersecurity assessment needs evidence of interpretation, prioritization, communication, and judgment, which a finished document alone rarely carries.

Academic integrity policies and AI detection tools address a different part of the problem. They may help institutions respond to misconduct, but do not show which evidence a student used, why a hypothesis changed, or whether the student can defend a recommendation. Responsible AI guidance adds needed attention to human oversight, privacy, transparency, and risk management (Autio et al., 2024; Miao & Holmes, 2023; Tabassi, 2023; U.S. Department of Education, 2023). Faculty still need assignment-level checks that connect those governance principles to evidence of learning.

The scope is intentionally narrow. The model is an assignment-level review whose conditions can also inform course and program governance. The paper focuses on adult and postsecondary cybersecurity education, meaning college, university, and professional or continuing-education settings rather than secondary schooling, although several design principles may transfer to other applied computing courses. It addresses assignments where students may use AI as part of the work, where AI may be embedded into a course activity, or where faculty must decide whether a proposed AI-supported assessment remains valid for grading. It does not evaluate a specific AI system, test student learning outcomes, or claim that the model has been empirically validated.

The four conditions are not chosen at random. Each condition derives from the literatures gathered above. Instructional purpose follows from evidence-centered design and authentic assessment, which tie a task to the competence it should reveal (Mislevy et al., 2003; Sotiriadou et al., 2020). Faculty judgment follows from the treatment of response processes and trace evidence as observable indicators an instructor must be able to interpret (Messick, 1995; Gašević et al., 2015). Assessment integrity follows from validity theory and from the distinction between assisted and unassisted performance that generative AI now forces (Messick, 1995; Pearce & Chiavaroli, 2023). Governance support follows from responsible AI guidance on oversight, disclosure, and data protection (Autio et al., 2024; Miao & Holmes, 2023; Tabassi, 2023; U.S. Department of Education, 2023). The conditions operate as interdependent layers rather than independently selected principles. Appendix A maps each to its foundation, the assignment-level gap it leaves, and the model's response.

5. A SIGNAL-PRESERVATION REVIEW MODEL

The model has four readiness conditions. Each condition asks a different question about evidence. Instructional purpose asks what the AI-supported task is meant to reveal. Faculty judgment asks whether instructors can see and assess the student's process. Assessment integrity asks whether what is visible is enough to support a fair inference about competence. Governance support asks whether the institution has workable rules for tools, data, disclosure, records, and review.

The conditions work together. A course can have a thoughtful AI policy and still use weak assignments. A faculty member can design strong process artifacts and still lack institutional guidance for data handling or approved tools. A model that focuses only on academic honesty misses faculty workload and validity. A model that focuses only on innovation misses the evidence needed for grading.

Table 1 presents the four conditions in a compact review model. It is intentionally brief because the purpose is to direct faculty attention to evidence, not to prescribe a universal assignment template. Each condition also implies a design action: naming the judgment assessed, selecting a manageable artifact, verifying the evidence supports the competence claim, and setting allowable AI use, disclosure, and records.

Condition	Guiding question	Evidence to preserve
Instructional purpose	What cybersecurity judgment is the task meant to develop?	Evidence selection, hypothesis formation, risk prioritization, framework use, or communication judgment.
Faculty judgment	Can faculty evaluate the process evidence as well as the final product?	Rubric language, examples, AI-use boundaries, and manageable review artifacts.
Assessment integrity	Does the work still support a fair inference about student competence?	Process artifacts, oral or written defense, revision trail, and comparison between process and product.
Governance support	Are tools, data, disclosure, documentation, and review expectations clear for cybersecurity courses?	Approved tools, data rules, disclosure language, recordkeeping expectations, and faculty authority.

Table 1. Signal-Preservation Review Model

5.1 Instructional Purpose

The first condition is instructional purpose. Before AI is added to a task, the instructor should be able to name the cybersecurity judgment the task is supposed to develop. That sounds obvious, but it is where many AI assignments become vague. Asking students to use AI to write a threat brief does not identify what they should learn; asking them to compare an AI-generated brief against packet evidence, revise it, and defend two rejected hypotheses is more precise.

Purpose determines the learning signal. In malware triage, the signal might be how students separate suspicious behavior from ordinary system noise; in a governance task, how they translate a broad policy into a control choice under constraints; in threat intelligence, confidence discipline, source evaluation, and attention to adversary intent. AI may assist each task, but the student must still produce evidence tied to the intended judgment.

A simple design rule follows. AI should be assigned to a function, not to the whole task. It may generate alternative explanations, produce a first-pass report structure, challenge a remediation plan, translate a technical finding for an executive audience, or identify missing evidence. Each function has a different assessment implication. If AI drafts the entire report, the instructor needs stronger companion evidence. If AI critiques a student draft, the student's response to the critique may become useful assessment evidence.

5.2 Faculty Judgment

The second condition is faculty judgment. AI-supported assessment asks faculty to evaluate a different evidentiary package than in the past. The final report remains useful but is no longer enough. Faculty need rubrics that reward evidence selection, uncertainty, sequencing, tool interpretation, and defense of judgment. They also need examples of acceptable and unacceptable AI use at the assignment level.

Faculty support should not be reduced to AI training. A faculty member need not become an AI engineer to ask better assessment questions; the harder work is curricular. What artifacts should students submit? Which parts may be AI-assisted, and which must be done without AI? How much process evidence is enough before grading becomes unmanageable? These are design decisions, and they require time, examples, and program-level consistency.

Faculty also need a vocabulary for partial AI use. A blanket statement such as "AI is permitted" creates more confusion than clarity. The assignment should say whether AI may be used for brainstorming, command explanation, grammar, report formatting, counterargument generation, or code review. It should also say where AI may not be used, such as fabricating evidence, inventing logs, replacing screenshots, or generating a reflection that claims firsthand learning. Without that specificity, faculty evaluation becomes inconsistent and student behavior becomes harder to interpret.

5.3 Assessment Integrity

The third condition is assessment integrity. Here, the word integrity refers to the strength of the inference, as well as to the honesty of the student. A cybersecurity assignment has assessment integrity when the evidence submitted is sufficient to support the claim being made about student competence. AI threatens that integrity when it breaks the link between the work product and the student's reasoning.

Signal-preserving assessment usually requires more than one artifact. For a threat analysis assignment, students might submit an evidence table, a hypothesis list, and a short final brief; for incident response, a decision log, an event timeline, and a post-action explanation. These companion artifacts need not be long. They have value because they let the instructor compare the process against the product. As a design heuristic, an instructor identifies the smallest set of artifacts that exposes the central judgment, often one process artifact with the final product, adding more only where it stays hidden. Preserved evidence should be proportionate to the outcome, judged for its analytic content, not its polish, and available in more than one form where feasible, so documentation burden does not become an unintended part of what the assessment measures.

Oral or dialogic components can strengthen this design, but they should be used carefully. A short defense of reasoning can reveal whether the student understands the work. It can also punish students who are anxious, have language barriers, or lack experience with professional questioning unless expectations are clear. Dialogic assessment works best when it is structured. A small number of predictable questions, a rubric for evaluating reasoning, and a record of the evidence being discussed can support fairness.

Assessment integrity also requires instructors to decide when AI output itself becomes evidence. If students are asked to critique an AI-generated risk memo, the generated memo is part of the task. If students use AI to improve prose after completing the analysis, the AI output may be less important than the student's marked revisions. Instructors should identify that distinction before the assignment begins.

5.4 Governance Support

The fourth condition is governance support. Cybersecurity courses add a layer of risk that many AI policies handle poorly. Students may work with logs, vulnerability descriptions, simulated incident data, infrastructure diagrams, or organization-like scenarios. Even when the data are synthetic, the course should teach habits that transfer to professional environments where data handling rules matter.

Responsible AI guidance, as cited above, consistently emphasizes human oversight, privacy, transparency, and risk management. For cybersecurity education, those principles need assignment-level translation. Students should know which tools are approved, what data may be entered, what must remain inside the learning environment, how AI use should be disclosed, and what records must be retained. Faculty should know what they may require and what support exists when students violate the rules.

Governance should also protect faculty authority. AI may help students practice analysis and may help faculty generate examples, but final judgments about student learning require accountable human review. The governance record should therefore connect the allowed AI function to the preserved evidence. If the policy allows AI-assisted drafting, the assignment should preserve the evidence that the draft was based on student analysis. If the policy allows AI critique, the student should document which suggestions were accepted and why.

6. APPLICATION TO COMMON CYBERSECURITY ASSIGNMENTS

The model becomes useful when it changes assignment design. In practice, applying the four conditions means naming the judgment being assessed, identifying the evidence that would demonstrate it, and determining which of it AI could obscure. The instructor then assigns AI only to functions that leave essential evidence intact, with companion evidence where it would not (Pearce & Chiavaroli, 2023), and before adopting the design, confirms that it stays gradable within a reasonable workload and meets the validity and governance conditions. The following examples show how the four conditions alter ordinary

cybersecurity tasks. The three cases are hypothetical design illustrations rather than evaluated interventions.

In a threat analysis assignment, a weak AI-supported design asks students to use AI to summarize suspicious network activity and submit a short brief. The final product may look professional while revealing little about the student's analytic judgment. A stronger design has students select three artifacts, write a hypothesis for each, identify one rejected explanation, and assign a confidence rating. AI may then generate alternative hypotheses or challenge the draft. The final submission includes the threat brief, the evidence table, the AI critique, and a short response on which AI suggestions were accepted or rejected, so the learning signal spans evidence selection, critique, and revision.

In an incident response reporting assignment, a weak design asks for a polished after-action report, which AI can largely supply. A stronger design has students keep a decision log during the scenario, each entry recording the observation, the decision, the alternatives, the expected risk, and what evidence would change it. AI may improve the executive summary afterward, but the student marks which sections were AI-assisted. The grade then rests on the decision log, event timeline, containment rationale, and final report, so the report no longer carries the assessment alone.

At the program level, a weak AI policy says students may use AI when the instructor allows it and must disclose use, which meets a minimal administrative need but does little for assessment. A stronger policy defines cybersecurity-specific data categories, from public and classroom-simulation data to restricted organizational and live operational data, and connects them to approved AI uses. It also requires each AI-supported assignment to identify the preserved learning signal, the AI functions allowed, the evidence collected, and the faculty review process, becoming an assessment design aid with usable requirements. The examples are condensed in Table 2.

Task	Compression risk	Signal-preserving revision
Threat analysis	AI produces a plausible narrative without showing which artifacts mattered.	Require an evidence table, rejected hypotheses, AI critique, and a response to the critique.
Incident response	AI organizes a polished after-action report while hiding decision sequence.	Require a decision log, event timeline, containment rationale, and marked AI-assisted sections.
Program AI policy	Policy permits AI use while leaving assignments without evidence requirements.	Require each assignment to name the signal preserved, the AI functions allowed, and the review process.

Table 2. Signal-Preserving Revisions for Common Cybersecurity Assignments

7. DISCUSSION

The model's value is that it moves the AI discussion closer to teaching. An instructor does not need a grand theory of AI adoption to revise a weak assignment, only to know which evidence must remain visible, which AI functions are acceptable, which artifacts will be graded, and which data cannot leave the learning environment. Those decisions are local, but they should not be improvised every week by individual faculty members working alone.

Learning signal compression also explains why some familiar responses feel unsatisfying. Banning AI can be reasonable when the goal is unaided recall, but a blanket ban may push use underground and leave faculty dependent on detection. Permitting AI without redesign accepts the tool while leaving the evidence thin. A signal-preservation approach allows AI when the task can preserve the evidence needed for assessment and restricts AI when that evidence would be erased.

The approach also respects the professional reality of cybersecurity. Students will work where AI-supported tools are increasingly present, and they need to use them without becoming passive consumers of generated explanations. A course that requires students to test AI output against evidence, defend corrections, and document uncertainty may teach a more realistic form of professional

judgment than a course that treats AI only as a cheating threat. The assessment, however, must be designed so that judgment remains visible.

Program leaders have a role here. Faculty cannot solve the data, tool approval, privacy, accessibility, and recordkeeping problems alone. Governance that stays at broad principles leaves too much burden on instructors, while governance that enters the assignment workflow supports good teaching. A curriculum committee should ask more than whether an AI policy exists. The stronger question is whether courses can reliably show what students actually did.

8. LIMITATIONS AND FUTURE RESEARCH

The proposal remains conceptual, which is its main limitation. It has not been tested through a controlled classroom study, inter-rater reliability analysis, or longitudinal evaluation of outcomes, and the assignment applications are design cases that do not show the model improves performance. Future research should test whether signal-preserving assignments improve the quality of student reasoning, reduce grading ambiguity, or change faculty confidence in AI-supported assessment.

Several research questions follow. Which process artifacts provide the strongest validity evidence? How much documentation can be required before grading becomes impractical? Which governance controls matter most for simulated cybersecurity data? These questions are empirical and should be studied across different course levels, institutional types, and student populations.

The model also needs usability testing with faculty. A review model that reads well on paper may fail in a curriculum meeting if it takes too long, uses language faculty do not share, or demands evidence instructors cannot realistically grade. The next step is a small design study in which cybersecurity faculty revise existing assignments using the model, document the changes, and evaluate whether the revised assignments preserve clearer evidence of learning.

9. CONCLUSION

Generative AI has not lowered the demand for judgment in cybersecurity work. It has made the record of that judgment easier to lose. An assignment that reads well can still leave a grader unable to see what the student noticed, ruled out, or defended.

A program that wants AI in the curriculum without weakening assessment has to decide, task by task, which reasoning stays visible and which tool functions may touch it. Those are design decisions, and they are most useful when faculty make them before the assignment runs rather than after a submission raises doubt.

The contribution is twofold. Conceptually, learning signal compression names how generative AI weakens the evidence behind a competence judgment, distinct from plagiarism and from validity in general. Practically, the four-condition model helps faculty preserve that evidence at assignment design rather than at detection. What remains unvalidated is whether such assignments improve student reasoning or grading reliability; the next step is a classroom study against real assignments and grading load.

10. REFERENCES

- Ajjawi, R., Tai, J., Dollinger, M., Dawson, P., Boud, D., & Bearman, M. (2024). From authentic assessment to authenticity in assessment: Broadening perspectives. *Assessment & Evaluation in Higher Education*, 49(4), 499-510. <https://doi.org/10.1080/02602938.2023.2271193>
- Autio, C., Schwartz, R., Dunietz, J., Jain, S., Stanley, M., Tabassi, E., Hall, P., & Roberts, K. (2024). Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile (NIST AI 600-1). National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.AI.600-1>
- Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, 59(1), 64-71. <https://doi.org/10.1007/s11528-014-0822-x>

- Glas, M., Vielberth, M., & Pernul, G. (2023). Train as you fight: Evaluating authentic cybersecurity training in cyber ranges. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (pp. 1-19). Association for Computing Machinery. <https://doi.org/10.1145/3544548.3581046>
- Joint Task Force on Cybersecurity Education. (2017). Cybersecurity curricula 2017: Curriculum guidelines for post-secondary degree programs in cybersecurity. Association for Computing Machinery, IEEE Computer Society, Association for Information Systems Special Interest Group on Security, and International Federation for Information Processing. <https://doi.org/10.1145/3184594>
- Kaldaras, L., Akaeze, H. O., & Reckase, M. D. (2024). Developing valid assessments in the era of generative artificial intelligence. *Frontiers in Education*, 9, Article 1399377. <https://doi.org/10.3389/feduc.2024.1399377>
- Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns*, 4(7), Article 100779. <https://doi.org/10.1016/j.patter.2023.100779>
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749. <https://doi.org/10.1037/0003-066X.50.9.741>
- Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO. <https://doi.org/10.54675/EWZM9535>
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1(1), 3-62. https://doi.org/10.1207/S15366359MEA0101_02
- MITRE. (n.d.). MITRE ATT&CK. <https://attack.mitre.org/>
- National Institute of Standards and Technology. (2024). The NIST Cybersecurity Framework (CSF) 2.0 (NIST CSWP 29). <https://doi.org/10.6028/NIST.CSWP.29>
- Pearce, J., & Chiavaroli, N. (2023). Rethinking assessment in response to generative artificial intelligence. *Medical Education*, 57(10), 889-891. <https://doi.org/10.1111/medu.15092>
- Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment. *Journal of University Teaching and Learning Practice*, 21(6). <https://doi.org/10.53761/q3azde36>
- Petersen, R., Santos, D., Wetzels, K., Smith, M., & Witte, G. (2020). Workforce Framework for Cybersecurity (NICE Framework) (NIST Special Publication 800-181 Revision 1). National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.SP.800-181r1>
- Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, 20(9), 676-688. <https://doi.org/10.1016/j.tics.2016.07.002>
- Sotiriadou, P., Logan, D., Daly, A., & Guest, R. (2020). The role of authentic assessment to preserve academic integrity and promote skill development and employability. *Studies in Higher Education*, 45(11), 2132-2148. <https://doi.org/10.1080/03075079.2019.1582015>
- Tabassi, E. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1). National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.AI.100-1>
- U.S. Department of Education, Office of Educational Technology. (2023). Artificial intelligence and the future of teaching and learning: Insights and recommendations. <https://www.ed.gov/media/document/ai-reportpdf-43861.pdf>

APPENDIX A

Foundations of the Signal-Preservation Model

Foundation	Contribution	Assignment-level gap	Model response
Evidence-centered design (Mislevy et al., 2003)	Links claim, evidence, and task	Does not address AI obscuring response evidence	Identifies the evidence that must remain observable
Validity theory (Messick, 1995)	Warrants an inference from performance	States the validity problem in general terms	Names the AI-mediated loss mechanism
AI Assessment Scale (Perkins et al., 2024)	Defines permissible AI involvement	Does not specify which cybersecurity evidence must stay visible	Ties each AI function to a preserved learning signal
Authentic assessment, CSEC2017, and NICE (Sotiriadou et al., 2020; Joint Task Force on Cybersecurity Education, 2017; Petersen et al., 2020)	Defines situated competence	Does not prescribe AI-era evidence preservation	Operationalizes evidence around cybersecurity judgment
Responsible AI guidance (Autio et al., 2024; Miao & Holmes, 2023; Tabassi, 2023; U.S. Department of Education, 2023)	Governance, transparency, and oversight	Mostly institutional or system level	Translates governance into assignment requirements