Issues in Information Systems
Volume 27, Issue 2, pp. 235-248, 2026

DOI: https://doi.org/10.48009/2_iis_2026_117

AI breaks the proxy: A framework for transforming IS pedagogy from execution to judgment

Bryan Marshall, *Georgia College & State University*, *bryan.marshall@gcsu.edu*
Peter Cardon, *University of Southern California, cardon@marshall.usc.edu*
Brad Fowler, *Georgia College & State University, brad.fowler@gcsu.edu*

Abstract

Generative AI has broken the execution proxies that IS education has relied on for decades. When students can produce correct SQL queries, functional web applications, and polished dashboards in minutes, execution no longer serves as evidence of understanding. This paper introduces the Artifact-Judgment Model (AJM), an assessment framework that reverses the traditional relationship: the artifact is given, the judgment is assessed. Students receive AI-generated work products and demonstrate understanding through evaluation, prediction, and trade-off reasoning rather than execution. The AJM is AI-proof by design — because assessment starts from AI output, using AI to complete it would simply reproduce the errors the assessment is designed to catch. The model is grounded in a three-tier framework analyzing how AI transforms IS competencies and is applied across six core IS domains. It aligns with AACSB 2026 Standard 4.3's mandate for human judgment over tool mastery and ABET's emphasis on AI as a support for, not replacement of, human judgment. The proxy problem is not unique to IS; the AJM offers a generalizable response to the defining pedagogical challenge of the AI era. The objective of this paper is to develop the AJM and demonstrate how it operationalizes judgment-based assessment across IS domains. As a conceptual and design contribution, it does not report empirical participant data; we specify a concrete validation pilot as the immediate next step.

Keywords: Artifact-Judgment Model, IS Curriculum, Artificial Intelligence, AI-Proof Assessment, Competency-Based Education, Pedagogy

Introduction

The information systems discipline stands at an inflection point. Generative AI tools have compressed the execution time for many traditional IS tasks from hours to seconds. A database query that once required careful SQL crafting can be generated from natural language. A web application that demanded weeks of development can be prototyped in an afternoon. A requirements specification that took days of interviews and documentation can be drafted from a single meeting transcript. These represent a compression of the execution layer of IS work — the layer where curricula have historically invested the majority of instructional time. For decades, IS education assessed understanding through execution: we asked students to write SQL queries to prove they understood databases, hand-code HTML to prove they understood web design, build dashboards to prove they understood analytics, and draw UML diagrams to prove they understood systems thinking. These were effective proxies when execution required understanding. AI breaks these proxies by making execution trivial. A student can now produce a correct query, a functional website, or a polished dashboard without understanding any underlying principles. The proxy no longer

works, forcing educators to confront a question they have never had to ask: how do we assess understanding directly, without the execution scaffolding that made it easy to conflate doing with knowing?

This proxy problem is not unique to information systems — every discipline that historically assessed understanding through execution faces the same rupture. While this paper analyzes IS as the specific use case, the framework we propose represents a generalizable pedagogical response.

The IS model curriculum's previous major evolution — from IS 2002's course-based prescriptive model to IS 2020's competency-based framework — was fundamentally philosophical, shifting from standardizing inputs (courses) to standardizing outputs (competencies). The transformation we now propose is fundamentally pedagogical — changing how every competency is taught and assessed, not merely what competencies are defined. A philosophical shift can be addressed by rewriting curriculum documents. A pedagogical shift requires faculty to change how they teach, what they assign, and how they evaluate student work, every day, in every course.

The urgency is underscored by workforce data. Workers in occupations requiring AI fluency grew sevenfold from approximately 1 million in 2023 to 7 million in 2025, the fastest-growing skill category in U.S. job postings (Schreiber-Shearer, 2025). Jobs requiring AI skills offer an average wage premium of 56%, up from 25% the prior year (PwC, 2025). AI skills now outperform formal educational qualifications in immediate labor market returns (World Economic Forum, 2025).

This paper addresses four research questions:

1. What framework can guide IS educators in identifying which competencies are augmented, transformed, or made more critical by AI?
2. How should specific IS curriculum domains be redesigned for AI-augmented professional practice?
3. What pedagogical and assessment methods can replace the execution proxies that AI has rendered ineffective?
4. How do emerging accreditation standards support this pedagogical transformation?

This study sits at the intersection of three literatures: the IS model-curriculum tradition (Gorgone et al., 2002; Leidig & Salmela, 2020), the rapidly growing scholarship on generative AI and assessment in higher education (Xia et al., 2024; Kofinas et al., 2025), and emerging accreditation guidance on AI and human judgment (AACSB, 2026; ABET, 2025). The recent assessment scholarship establishes that AI compels a transformation of assessment but has largely advanced *defensive* or *authenticity-based* responses; as we show in Section 2.3, neither fully resolves the proxy problem. The contribution of this paper is to move from diagnosis to design: a tier-based analysis of which competencies AI augments, transforms, or makes more critical, and a concrete, AI-resilient assessment model—the Artifact-Judgment Model—that other disciplines can adopt.

Background and literature review

IS curriculum evolution and the pattern of additive response

The IS model curriculum has evolved through several revisions, each reflecting the technological realities of its era. IS 2002 provided a course-based prescriptive framework defining specific courses such as Data Management, Networking, Programming, and Systems Analysis (Gorgone et al., 2002). IS 2020 shifted to a competency-based approach, identifying 19 competency areas grouped into six realms and applying Bloom's taxonomy to define proficiency levels (Leidig & Salmela, 2020). This was a significant philosophical evolution — from standardizing inputs (courses) to standardizing outputs (competencies).

The historical response pattern is instructive: each past disruption was addressed by adding content. This additive approach works when disruption creates new knowledge requiring new curricular space. AI represents a categorically different disruption. Rather than adding a topic, AI changes how every existing topic should be taught. Database management is not unchanged-plus-AI. The additive approach fails

because it treats AI as content to be learned rather than a force that reshapes how all content is taught and assessed.

Critically, both IS 2002 and IS 2020 retained an implicit assumption: that competency development requires sustained manual practice, and that assessment relies on observing execution. This assumption was reasonable before generative AI demonstrated that execution can occur without understanding.

AI's impact on knowledge work and education

Gartner estimates that over 80% of enterprises will have deployed GenAI-enabled applications by 2026 and predicts that 50% of organizations will require "AI-free" skills assessments due to concerns about critical-thinking atrophy (Gartner, n.d.). The McKinsey Global Institute estimates generative AI could automate 60-70% of tasks currently occupying knowledge workers' time, with impact concentrated in precisely the areas where IS curricula invest the most instructional time: writing code, creating documentation, building reports, and producing analysis artifacts (McKinsey, 2023).

This creates a paradox for IS educators: students must become proficient with AI tools to remain competitive, yet must also develop the foundational knowledge to use AI tools effectively and recognize when outputs are incorrect. A graduate who cannot use AI tools will be inefficient. A graduate who uses AI tools without understanding the domain will be unreliable. The curriculum must develop both capabilities simultaneously, and the pedagogy must be designed so that AI tool use itself becomes the mechanism through which conceptual understanding is developed and assessed.

The institutional response has been largely defensive. AACSB found that integrity concerns are the leading barrier preventing faculty from integrating AI into teaching (AACSB, 2025). This defensive posture reflects the same conflation of execution with understanding that defines the proxy problem. When assessment is designed around execution tasks AI can perform, any AI use looks like cheating. When assessment is redesigned around judgment AI cannot provide, the cheating question dissolves.

Institutions are beginning to recognize this distinction. Harland (2026) documents business schools moving beyond the defensive stage toward treating AI as a pedagogical partner, with faculty grading students on their ability to evaluate, revise, or reject AI outputs rather than on original execution. Askari (2025) argues that integrity violations linked to AI-assisted assessments are fundamentally stories about institutional choices — what to measure, how to assess — concluding that "the greatest risk business schools face in the age of AI isn't that students will cheat. It's that institutions will outsource ethical reasoning to systems not designed for it."

These institutional accounts are increasingly corroborated by peer-reviewed scholarship. A scoping review of 32 empirical studies concludes that generative AI compels a transformation of assessment toward designs that cultivate self-regulated learning and integrity rather than detection (Xia et al., 2024); the review's cross-disciplinary breadth is a strength, though it also cautions that the empirical base remains early and heterogeneous. We note, too, that several of the workforce and enterprise figures cited above derive from industry reports (Schreiber-Shearer, 2025; PwC, 2025; World Economic Forum, 2025; Gartner, n.d.; McKinsey, 2023): timely and large-scale, but not peer-reviewed, and therefore treated here as indicative context rather than confirmed effect sizes.

Approaches to AI-resilient assessment

The literature offers several responses to the integrity threat AI poses to assessment, each with characteristic strengths and limitations. **Authentic, work-integrated assessment** raises real-world relevance and student motivation; however, Kofinas et al. (2025) demonstrate that authenticity alone does not preserve integrity, because students can have generative AI produce the reflections, case analyses, and project artifacts that authentic tasks require. **Secure or invigilated assessment**, including a return to proctored and oral examinations, is robust against unauthorized AI use but sacrifices authenticity and scales poorly, and oral defenses are resource-intensive. **Process- and product-based approaches** assess the trajectory of a student's work and AI use, not only the final artifact (Awadallah Alkouk & Khlaif, 2024); their strength is that they engage AI use directly, but they depend on observing process and are harder to standardize across

sections. **Guidance frameworks** that teach students when AI use is and is not appropriate (Firth et al., 2024) address the integrity question at the level of policy and judgment, yet they do not by themselves redesign what is assessed.

The Artifact-Judgment Model developed in Section 5 occupies a distinct position. Rather than making the task harder for AI (secure assessment), more realistic (authentic assessment), or more transparent (process assessment), it starts from AI output and grades the human judgment applied to it—remaining authentic and AI-resilient at once. We revisit this comparison once the model is specified.

Theoretical framework: AI-augmented capability tiers

We propose a three-tier framework for analyzing how AI transforms IS competencies. The framework emerges from recognition that much of what IS curricula taught as "skills" were actually execution proxies for deeper understanding. Writing SQL stood in for understanding the relational model. Coding HTML stood in for understanding design. Building dashboards stood in for understanding analysis. AI makes this conflation visible by breaking it. When AI handles execution, we must distinguish between three layers of capability: the understanding that enables judgment (Tier 1), the execution AI now performs (Tier 2), and the new evaluative skill of directing and auditing AI output (Tier 3).

Tier 1: Conceptual foundations

Tier 1 encompasses knowledge and mental models AI cannot replace — the "why" of IS practice. AI makes foundational knowledge *more* important, not less: understanding the relational model is precisely what enables a professional to recognize when AI-generated SQL is wrong, inefficient, or answering the wrong question. Without Tier 1 understanding, a professional cannot evaluate AI output; they can only accept it.

Examples include data modeling and normalization theory, business process understanding, and systems thinking. These develop slowly through conceptual engagement, transfer across tools and platforms, and provide the judgment layer for supervising AI.

Pedagogical implication: Teach for understanding, not memorization. The time freed by compressing execution drill should be reinvested in conceptual depth, not breadth.

Tier 2: AI-executable tasks

Tier 2 encompasses skills traditionally requiring substantial practice that AI now performs rapidly — the "how" that can be delegated when the user has sufficient Tier 1 understanding. The pedagogical challenge is that these tasks previously served as the primary mechanism for *building* understanding. When we compress Tier 2 execution, we must replace this learning mechanism, not just eliminate it.

Examples include writing SQL from specifications, generating HTML/CSS/JavaScript, and drafting requirements documents.

Pedagogical implication: Shift from teaching execution to teaching effective AI direction. Students should still engage with Tier 2 tasks — not by performing them manually, but by reading, evaluating, and refining AI output. Learning happens through evaluation, not execution.

Tier 3: Human-AI integration

Tier 3 encompasses the emerging capability to effectively direct, evaluate, and govern AI systems. This is not "prompt engineering" in the narrow sense but a broader professional capability: knowing what to ask AI to do, evaluating whether AI did it correctly, recognizing the boundaries of AI reliability, designing appropriate workflows, and governing AI use at an organizational level.

Examples include evaluating AI-generated schemas for normalization errors, auditing AI configurations for security vulnerabilities, and detecting bias and hallucination in AI-synthesized reports.

Pedagogical implication: Tier 3 must be explicitly taught through practice with real AI tools, including deliberate exposure to AI failures. The most effective pedagogy involves students working with AI output that is plausible but flawed — the IS equivalent of medical education's emphasis on differential diagnosis.

Framework dynamics

The three-tier framework is not rigid — boundaries shift as AI capabilities evolve. Its value lies in providing a lens for analyzing how instructional time should be reallocated. The framework reveals a critical asymmetry: while Tier 2 tasks can be compressed quickly, Tier 1 understanding and Tier 3 evaluation skills take just as long to develop as they always have. Curriculum redesign is therefore a matter of reallocating time, not removing content.

Applying the framework: domain analysis

Competency mapping

Table 1 summarizes how IS 2020 competency areas map to the three tiers.

Table 1. IS 2020 Competency Areas Mapped to The Three Capability Tiers.

IS 2020 Competency Area	Tier 1 (Foundations)	Tier 2 (AI-Executable)	Tier 3 (Human-AI Integration)
Data/Information Management	Data modeling, normalization, query optimization	SQL writing, basic ETL	Evaluating AI-generated schemas and queries
IT Infrastructure	Network architecture, security principles	Configuration syntax, documentation	Auditing AI-generated configs, security monitoring
Application Development	Software architecture, design patterns	Code generation, UI implementation	Code review of AI output, AI-augmented debugging
IS Project Management	Stakeholder management, risk assessment	Status reporting, estimation	Managing AI project lifecycles
Analytics & BI	Analytical thinking, data storytelling	Dashboard creation, metric calculation	Validating AI insights, augmented analytics governance
Systems Analysis & Design	Systems thinking, requirements analysis	Requirements generation, use-case modeling	Designing systems with AI components
Secure Computing	Security principles, threat modeling	Configuration, documentation	AI-specific threats (prompt injection, model poisoning)
Ethics & Implications	Ethical reasoning frameworks	N/A	AI governance, responsible AI, bias detection

Detailed domain analyses

The following sections analyze six core curriculum domains. To avoid duplication, each domain is treated once: §4.2 gives the proxy-problem rationale and representative assessments, while the consolidated per-domain mapping of what students receive, do, and are graded on is formalized in Table 2 (Section 5). Database management, web design, and systems analysis are shown with worked assessment examples; agile PM, networking, and BI are summarized in a single concise rationale each.

Database management

The proxy problem: SQL query writing was the proxy for understanding the relational model. AI now generates syntactically correct SQL from natural language in seconds, so the distinguishing skill is reading, evaluating, and reasoning about SQL rather than producing it: examining a query to judge what a join accomplishes, whether a subquery is redundant, why it returns incorrect results, and whether it answers the

business question that prompted it. That judgment rests on deep understanding of normalization, query optimization, and indexing—a student who grasps third normal form spots the update anomaly an AI-generated schema introduces, and one who reads execution plans catches the full table scan where an indexed approach belongs.
Assessment examples: - Evaluation: "AI generated this three-table join. The query runs without errors but returns incorrect results. Identify the problem and explain why the join type matters." - Prediction: "Here is a query on a table with 2 million rows and no indexes. Before running it, predict performance. Now add an index — predict how performance changes and why." - Trade-off: "Here are two schema designs: fully normalized versus denormalized for reporting. Which would you recommend for 500 orders/day with hourly executive reports? Justify your reasoning."

Agile project management
The proxy problem: Scrum ceremony execution and administrative tasks were proxies for stakeholder judgment and leadership through ambiguity. AI now automates user-story generation, story-point estimation, sprint reporting, and stakeholder communication — exposing the distinction between project administrators (who maintain artifacts) and project leaders (who navigate organizational complexity under uncertainty). Tier 1 deepens around stakeholder navigation and risk; Tier 3 adds quality governance for AI-generated deliverables and recalibration when AI coding assistants alter velocity. **Representative assessment:** "AI generated a sprint plan for a team of four. Identify which stories are too vague, which estimates are unrealistic, and what risks the plan ignores."

Networking and infrastructure
The proxy problem: Manual subnet calculation and configuration syntax served as proxies for understanding network architecture and security principles. AI now generates complete VPC configurations with subnets, route tables, security groups, and NACLs from natural-language descriptions in seconds, exposing the gap between students who memorized procedures and those who understand how networks work. Tier 1 maintains and deepens around DNS, subnetting, layered security, and architectural resilience; Tier 3 adds security evaluation of AI-generated configurations — an AI-generated firewall rule that opens port 22 to 0.0.0.0/0 looks like a valid answer to "set up SSH access" but is recognized as an invitation to brute-force attack by anyone with security fundamentals. **Representative assessment:** "AI generated this VPC configuration with security groups, route tables, and NACLs. Identify the security vulnerabilities and explain what each exposes."

Web design and application development
The proxy problem: Hand-coding was the proxy for understanding software design. AI now generates complete, functional, responsive websites from natural language in under a minute. Tools like Claude, Cursor, and Bolt produce full-stack applications from a paragraph of requirements. The traditional model of dedicating weeks to hand-writing HTML, debugging CSS layouts, or scaffolding boilerplate allocates substantial time to competencies AI executes in seconds. What AI cannot do is determine what should be built, evaluate whether results meet user needs, or identify when architecturally unsound decisions will create technical debt at scale.
Assessment examples: - Evaluation: "AI generated this web application from a design brief. Run a Lighthouse audit, review the component structure, and identify accessibility violations, performance bottlenecks, and architectural decisions that will not scale." - Prediction: "Here is a React component tree with state at three levels. Before running it, predict what happens when a user updates their profile. Which components re-render and why?" - Trade-off: "A client needs an internal dashboard. Compare server-rendered with minimal JavaScript versus a React SPA with REST API. One developer, 50 internal users. Which approach and why?"

Business intelligence and analytics
The proxy problem: Dashboard creation was the proxy for analytical thinking. AI now builds dashboards, generates DAX measures, and produces narrative summaries from natural language — the tool proficiency BI courses spent weeks developing is no longer the scarce skill. What remains scarce is the ability to look at data and ask the question no one thought to ask, to recognize when a metric that looks good actually masks a serious problem, and to translate findings into recommendations that account for organizational context the data does not contain. Tier 1 deepens around analytical framing and professional skepticism; Tier 3 adds validation of AI-identified anomalies and governance for natural-language query interfaces. *Representative assessment:* "AI flagged a 30% revenue spike in Q3 as a key insight. Examine the data. Is this a genuine trend, seasonal pattern, or artifact from a one-time bulk order?"

Systems analysis and business analysis
The proxy problem: Drawing UML diagrams, writing formal use cases, and producing requirements documents were proxies for systems thinking and the ability to understand what an organization actually needs. AI now generates use cases, sequence diagrams, data flow diagrams, and requirements specifications from natural language in seconds — exposing that diagramming proficiency was never the same thing as analytical reasoning about complex organizational problems.
Assessment examples: - Evaluation: "AI generated a requirements specification from a stakeholder transcript. Compare it to the original and identify three requirements AI inferred but the stakeholder never stated, and two needs the stakeholder expressed that AI missed." - Prediction: "Here is a proposed system change automating purchase order approvals. Predict how it affects the three departments currently in the process. Which roles gain or lose authority?" - Trade-off: "A university wants to replace its legacy registration system. Compare an 18-month parallel migration versus a hard cutover during summer. The registrar has four staff and no IT department. Which approach and why?"

The Artifact-Judgment Model (AJM)

The proxy problem across all six domains converges on a single question: if execution no longer proves understanding, what does? The solution is not banning AI or finding new proxies AI cannot yet perform. It is to stop assessing execution and assess understanding directly.
We propose the **Artifact-Judgment Model (AJM)**: the artifact is given, the judgment is assessed. Students receive AI-generated work products — queries, configurations, applications, dashboards, project plans, requirements documents — and demonstrate understanding through what they can see that AI cannot. The AJM reverses the traditional assessment relationship. Where the old model started from a blank screen and graded the artifact a student produced, the AJM starts from an AI-generated artifact and grades the judgment a student applies to it.

Three assessment methods
The model operationalizes through three assessment methods:
Evaluation-as-learning. Replace execution drills with critical evaluation. Instead of "write 50 queries to internalize SQL," the new model is "read and evaluate 50 queries to internalize relational thinking." Students receive AI-generated output — schemas, configurations, code, dashboards, project plans — containing subtle errors, inefficiencies, or misalignments with business requirements. Diagnosis becomes the core learning activity, and understanding develops through critical evaluation rather than production.
Prediction before execution. Before running a query, students predict what it returns. Before deploying a configuration, they predict what it does. Before launching a dashboard, they predict what story the data will tell. Prediction forces mental-model construction — students cannot predict behavior without understanding mechanism — and is the assessment AI cannot shortcut, because the point is what the student understands, not what appears on screen.

Trade-off reasoning under constraints. Present students with two or more valid approaches and require them to argue for one given specific business context. There is no single right answer — the quality of reasoning determines the grade. This mirrors what IS professionals actually do: they rarely build from scratch; they choose between options under constraints that require judgment AI does not have.

This design resolves the limitation identified in Section 2.3. Authentic assessment fails when AI can complete the authentic task; secure assessment forgoes authenticity; process assessment depends on observing how students work. The AJM keeps the task authentic—students do what practitioners actually do, evaluating and deciding about AI output—while remaining resilient to AI use, because what is graded is the judgment AI cannot supply about its own work.

Framework formalization

Table 2 formalizes this approach across all six domains.

Table 2. The Artifact-Judgment Model Across Six IS domains.

Domain	What Students Receive	What Students Do	What Gets Graded
Database Management	AI-generated queries, schemas, execution plans	Evaluate SQL for correctness and performance; predict query behavior; choose between schema designs	Diagnostic reasoning; prediction accuracy; trade-off depth
Networking & Infrastructure	AI-generated configurations, architecture diagrams	Security audit; predict traffic under failure; recommend architectures for constraints	Vulnerability identification; behavioral predictions; architectural reasoning
Web Design & Development	AI-generated applications, Lighthouse results	Code review for accessibility, security, architecture; predict state behavior	Code review depth; behavior predictions; architectural trade-offs
Business Intelligence	AI-generated dashboards, anomaly reports	Validate insights as genuine or artifacts; predict data patterns; prioritize KPIs	Signal vs. noise distinction; data intuition; stakeholder communication
Agile Project Management	AI-generated sprint plans, risk assessments	Evaluate feasibility and risk; predict downstream effects; navigate stakeholder conflicts	Plan critique quality; impact predictions; stakeholder communication
Systems & Business Analysis	AI-generated requirements, system models	Compare against stakeholder transcripts for gaps; predict organizational impact	Identification of AI fabrications; impact predictions; implementation reasoning

Each domain in Section 4 demonstrates how these methods apply to specific IS competencies.

Model properties

The AJM has several properties that distinguish it from traditional assessment.

AI-proof by design. This is the AJM's defining property. Because the assessment starts from AI output, there is nothing for AI to produce on the student's behalf — a student who asks AI to evaluate AI-generated code receives the same errors the assessment is designed to catch. AI cannot reliably identify its own failure modes. The value that gets graded is the human judgment AI cannot provide about its own output. Faculty who adopt the AJM report that the academic-integrity question simply disappears (Harland, 2026): not because students stop using AI, but because the assessment is designed so that AI use and cheating are no longer the same thing.

Robustness as AI improves. A natural objection is that this AI-proofing depends on a current limitation—AI's difficulty evaluating and debugging its own output—that more capable models may overcome. Three considerations bound this risk. First, much of what the AJM grades is not self-evaluation but judgment under context the artifact does not contain: which trade-off fits this organization's constraints, which

stakeholder need an AI-generated specification missed, whether a metric that looks good masks a real problem. These depend on information outside the model's input, not on better introspection. Second, even if a model could evaluate its own output, a student who delegates that judgment learns nothing—so the assessment still measures the student's understanding regardless of what AI can do. Third, the model is adaptive: as AI improves at catching obvious errors, faculty escalate the subtlety of curated artifacts and shift weight toward prediction and trade-off reasoning, the methods least reducible to pattern completion. The AJM is therefore best understood not as a permanent exploit of an AI weakness but as an assessment design whose validity survives improvement in the tools, because it grades the human contribution that remains scarce.

Assesses understanding directly. Rather than inferring understanding from execution ability, the AJM asks students to demonstrate it explicitly — through explanation, prediction, and reasoning. A student who cannot explain why an AI-generated query is inefficient does not understand query optimization, regardless of whether they could write an efficient query themselves. This directness produces higher-fidelity evidence of learning than execution proxies ever did.

Mirrors professional practice. In contemporary IS practice, professionals increasingly work with AI-generated artifacts rather than creating everything from scratch. The ability to evaluate, refine, and make decisions about AI output is rapidly becoming the core professional skill in every IS domain. The AJM develops and measures this skill directly.

Scales across proficiency levels. The same model works for introductory and advanced courses by adjusting the complexity and subtlety of artifacts. An introductory database course might present an AI-generated query with an obvious join error; an advanced course might present a query that is correct but performs poorly at scale due to a missing index on a non-obvious column. The assessment structure remains the same; the depth of understanding required increases.

Implications for practice

Accreditation alignment

The pedagogical transformation proposed here is not merely compatible with current accreditation standards — it is increasingly mandated by them.

AACSB Standard 4: Curriculum and Digital Agility. The 2026 Global Standards require curricula to cultivate "digital agility" and "responsible and ethical use of technology, emphasizing human judgment, critical evaluation, and appropriate application rather than mastery of specific tools or platforms" (AACSB, 2026). The Artifact-Judgment Model provides the pedagogical mechanism for achieving this.

AACSB Standard 5: Assurance of Learning. AACSB warns that using AI tools without proper pedagogical design can produce "performative signals" — metrics that simulate readiness without reflecting genuine learning (AACSB, 2026). The AJM provides an AI-resistant Assurance of Learning mechanism that measures true comprehension rather than AI-reproducible outputs.

AACSB Standard 9: Societal Impact. The 2026 standards require schools to integrate responsible AI and ethics throughout curricula (AACSB, 2026). Tier 1 conceptual foundations provide the ethical scaffolding accreditors require.

ABET. ABET affirms that "AI technologies are intended to support and enhance human judgement and interaction, not replace them" and that "AI must not replace human judgement in the continuous improvement process" (ABET, 2025). This mirrors the paper's argument precisely.

Table 3 makes these alignments concrete, mapping specific accreditation requirements to the AJM mechanism that satisfies each and the evidence it produces.

Table 3. Mapping Accreditation Requirements to Artifact-Judgment Model Mechanisms.

Standard / criterion	What it requires	AJM mechanism that satisfies it	Assessable evidence
AACSB 4.3 (Curriculum, digital agility)	Human judgment and critical evaluation over tool mastery	Evaluation-as-learning grades critique of AI output, not tool operation	Diagnostic write-ups identifying errors in AI-generated artifacts
AACSB 5 (Assurance of Learning)	Genuine learning measures, not "performative signals"	Prediction-before-execution cannot be shortcut by AI	Prediction-accuracy records tied to stated learning goals
AACSB 9 (Societal Impact)	Responsible-AI and ethics integration	Tier 3 governance tasks: bias, security, and fabrication audits	Student audits of AI output for bias and hallucination
AACSB 3 & 7 (Faculty sufficiency; teaching effectiveness)	Demonstrated faculty capacity and teaching impact	Assessment-redesign-first faculty development	Redesigned AJM assessments and shared rubrics
ABET (AI supports, not replaces, judgment)	AI must not replace human judgment	Trade-off reasoning grades student judgment under constraints	Justified recommendations chosen among valid alternatives

The convergence of these positions signals that the transformation from execution-based to judgment-based assessment is an emerging consensus across accrediting bodies spanning business education, computing, and engineering.

New competency areas

Several competency areas warrant increased emphasis. **AI Governance and Responsible AI** — including organizational policy development, the NIST AI Risk Management Framework (NIST, 2023), ISO/IEC 42001 standards, bias detection, and responsible AI principles — is elevated from peripheral coverage to an essential competency by AACSB Standard 9's requirement for societal impact integration (AACSB, 2026). **AI-Augmented Business Process Automation** — including agentic AI systems, workflow automation, and human-in-the-loop design — represents a new capability category current curricula do not address. **Human-AI Collaboration** — including calibrated trust frameworks, critical evaluation across risk levels, and collaborative workflow design — extends the general AI-literacy competencies identified by Long and Magerko (2020) into the IS-specific professional judgment that integrates AI capabilities in ways that amplify rather than diminish outcome quality.

Structural principles for curriculum redesign

1. **Compress Tier 2 deliberately.** Identify which execution tasks are being compressed, what learning they previously provided, and how that learning will be replaced.
2. **Deepen Tier 1 foundations.** Invest time freed from execution drill in deeper conceptual treatment.
3. **Add explicit Tier 3 instruction** in every course — not as a separate AI module but woven throughout.
4. **Integrate AI tools throughout** all coursework, mirroring professional practice.
5. **Emphasize judgment through deliberate error exposure.** Curate AI output with realistic, domain-specific errors for students to diagnose.

Implementation considerations

Faculty development. This transformation requires faculty to rethink courses they may have taught for decades. The AACSB 2026 Standards frame this as an accreditation requirement: Standard 3 (Faculty Sufficiency) requires demonstrated faculty capacity, and Standard 7 (Teaching Effectiveness) elevates teaching impact to a measurable requirement (AACSB, 2026). Faculty development programs should lead with assessment redesign — showing how the Artifact-Judgment Model resolves the integrity dilemma — rather than AI tool training. Faculty who see that redesigned assessment eliminates the cheating problem are more likely to embrace AI integration.

Tool access and equity. When AI is integrated throughout the curriculum, institutions must provide equitable access through institutional subscriptions rather than expecting students to bear costs individually. The AACSB 2026 framework emphasizes the democratization of AI education — shifting AI from elective to foundational and ensuring equitable access for all students (AACSB, 2026).

Toward empirical validation: an immediate next step

The AJM is presented here as a conceptual and design contribution; its claims about learning are not yet empirically tested. The immediate next phase, now being planned, is a multi-section pilot. In an introductory IS course, two sections will be taught with matched content but different assessment regimes—one execution-based, one AJM-based—with sections matched on prior GPA and a baseline pre-test. Outcomes will be measured with (a) a transfer task in an unseen domain scored with a shared analytic rubric, (b) a prediction task that AI cannot shortcut, and (c) a delayed post-test of conceptual understanding, alongside student and faculty experience surveys. Inter-rater reliability will be established for rubric scoring, and threats to validity—instructor effects, self-selection, and novelty—will be addressed through replication across terms and instructors. Even a single deployed AJM assessment, such as the AI-generated SQL audit described in Section 4, provides an early anchor of feasibility and student response. This design converts the empirical gap noted below into a concrete, near-term research program rather than an open-ended aspiration.

Limitations and future research

The three-tier framework and Artifact-Judgment Model have not yet been empirically tested through controlled implementation. The claims about learning effectiveness — that evaluation-based learning can replace execution-based learning, that prediction exercises build mental models as effectively as practice — remain to be validated through rigorous pedagogical research. Future research should conduct pilot studies comparing student outcomes in traditional versus AI-augmented sections across multiple institutions.

AI capabilities are advancing rapidly, and recommendations appropriate for 2026 may require revision as tools evolve. The framework itself is durable — distinguishing what AI can do, what humans must understand, and what the human-AI interface requires.

Implementation feasibility varies significantly by institution. A large research university with substantial technology budgets faces different challenges than a small teaching-focused institution with limited resources. The domain analyses draw primarily from the authors' experience and would benefit from multi-institutional validation.

The Artifact-Judgment Model's psychometric properties — reliability, validity, fairness — have not been established. Developing and validating assessment instruments for evaluation, prediction, and trade-off reasoning across IS domains is a substantial research agenda. Additional future research directions include longitudinal studies of student outcomes, employer satisfaction surveys, and investigation of whether evaluation-based learning develops the same depth of understanding as execution-based learning.

Conclusion

IS education built its pedagogy around execution proxies — writing queries, coding pages, building dashboards, configuring networks, drawing diagrams, running ceremonies. These proxies worked because execution required understanding. AI separates them: AI commoditizes execution, breaking the proxy and forcing a question IS education has never had to answer directly: do our students actually understand?

The three-tier framework provides a systematic approach for analyzing this transformation, and the Artifact-Judgment Model translates it into concrete assessment practice: the artifact is given, the judgment is assessed. The model is AI-resistant by design, mirrors professional practice, and assesses understanding directly rather than inferring it from execution ability.

Interpreted more broadly, the framework reframes what an IS program is for in the AI era. If execution is commoditized, the distinctive value a graduate brings is the judgment to direct, evaluate, and govern AI—precisely the Tier 1 understanding and Tier 3 evaluative capability the framework foregrounds and the AJM measures. A curriculum organized this way prepares students not to compete with AI on execution, a contest they will lose, but to do what AI-enabled workplaces increasingly demand: decide what to build, recognize when AI output is wrong or ill-suited to context, and take professional responsibility for outcomes. The three-tier framework gives faculty a shared language for reallocating instructional time toward these capabilities, and the AJM gives them a means of assessing—and therefore credibly certifying—that students have developed them. The contribution is thus not only a new assessment technique but a way of keeping the IS degree a meaningful signal of competence when the tasks it once certified can be automated.

This proxy problem is not confined to information systems — it is the defining pedagogical challenge of the AI era across higher education, and AACSB's 2026 Standards and ABET's AI position both mandate the shift from tool mastery to human judgment. The strategic choice facing IS programs is not between adopting AI and ignoring it. It is between transforming pedagogy to develop evaluative judgment and continuing to develop execution competencies whose professional half-life is shorter than the time required to complete a degree program.

Acknowledgment of Generative AI Use

This paper was developed with the assistance of generative AI tools (Claude, Anthropic). AI was used as a writing and editing aid for drafting, restructuring, and refining prose. All intellectual contributions — including the three-tier framework, the Artifact-Judgment Model, domain analyses, and all arguments presented — are the original work of the authors. The authors reviewed, validated, and take full responsibility for the originality, validity, and integrity of all content.

References

AACSB International. (2025). *GenAI adoption in business schools: Deans and faculty respond*. https://www.aacsb.edu/insights/reports/2025/genai-adoption-in-business-schools-deans-and-faculty-respond

AACSB International. (2026). *2026 global standards for business education*. (Exposure draft). https://www.aacsb.edu/-/media/documents/accreditation/2026/2026-global-standards.pdf?rev=c405d18578c24dddb929f88ac7834cd6

ABET. (2025). *ABET statement on artificial intelligence*. https://www.abet.org/accreditation/ai-policy/

Askari, M. (2025). The AI dilemma: When innovation outpaces integrity. *AACSB Insights.* https://www.aacsb.edu/insights/articles/2025/10/the-ai-dilemma-when-innovation-outpaces-integrity

Awadallah Alkouk, W., & Khlaif, Z. N. (2024). AI-resistant assessments in higher education: Practical insights from faculty training workshops. *Frontiers in Education*, 9, 1499495. https://doi.org/10.3389/feduc.2024.1499495

Firth, D. R., Derendinger, M., & Triche, J. (2024). Cheating better with ChatGPT: A framework for teaching students when to use ChatGPT and other generative AI bots. *Information Systems Education Journal*, 22(3), 47-60.

Gartner. (n.d.). Harness the power of democratized generative AI to transform your business [Video]. https://www.gartner.com/en/experts/top-tech-trends-unpacked-series/harness-democratized-generative-ai-transform-your-business

Gorgone, J. T., Davis, G. B., Valacich, J. S., Topi, H., Feinstein, D. L., & Longenecker, H. E. (2003). IS 2002: Model curriculum and guidelines for undergraduate degree programs in information systems. *Communications of the Association for Information Systems*, 11(1). https://doi.org/10.17705/1CAIS.01101

Harland, N. (2026). How AI is evolving in business schools. *AACSB Insights*. https://www.aacsb.edu/insights/articles/2026/02/how-ai-is-evolving-in-business-schools

Kofinas, A. K., Tsay, C. H.-H., & Pike, D. (2025). The impact of generative AI on academic integrity of authentic assessments within a higher education context. *British Journal of Educational Technology*, 56(6), 2522-2549. https://doi.org/10.1111/bjet.13585

Leidig, P. M., & Salmela, H. (2021). IS 2020: *A competency model for undergraduate programs in information systems*. Communications of the Association for Information Systems. https://doi.org/10.1145/3460863

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1-16. Association for Computing Machinery. https://doi.org/10.1145/3313831.3376727

McKinsey Global Institute. (2023*). The economic potential of generative AI: The next productivity frontier*. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier

National Institute of Standards and Technology. (2023). *AI risk management framework* (AI RMF 1.0) (NIST AI 100-1). U.S. Department of Commerce. https://doi.org/10.6028/NIST.AI.100-1

PwC. (2025). The fearless future: 2025 global AI jobs barometer. https://www.pwc.com/gx/en/issues/artificial-intelligence/job-barometer/2025/report.pdf

Schreiber-Shearer, N. (2025, December 17). AI skills demand in the U.S. job market. Gloat. https://gloat.com/blog/ai-skills-demand/

World Economic Forum. (2025). *Future of jobs report 2025*. https://www.weforum.org/publications/the-future-of-jobs-report-2025/

Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. F. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. *International Journal of*

Educational Technology in Higher Education, 21(1), 40. https://doi.org/10.1186/s41239-024-00468-z