Composition Pedagogy as AI-Native Coding: From Design Kit to Scholarly Framework

Daniel Plate[1*], James Hutson[2]

Lindenwood University, USA.

Abstract: This article advances a field-ready framework that reconceives first-year composition as AI-native coding, translating a complete "design kit" into scholarly method, evaluative protocol, and curriculum architecture. Background: Contemporary composition pedagogy emphasizes process, genre awareness, and collaborative revision; meanwhile, modern software practice operationalizes iteration through version control, test-driven development, and continuous integration. The uploaded kit demonstrates that these cultures are isomorphic: writing stages align with SDLC phases, and automated pipelines can lint prose, execute argument "tests," and publish artifacts with auditable histories. Approach: The study systematizes that kit into (1) a conceptual map that recasts authorship as orchestration and verification, (2) a pipeline specification that integrates rhetorical linters, claim-evidence checks, retrieval-grounded fact audits, and CI dashboards, and (3) an assessment regime that grades specification quality, revision discipline, and process transparency alongside argument strength and source integration. Significance of results: The framework yields inspectable process evidence that reduces adjudication ambiguity, raises floor quality on conventions through automation, and reallocates instructor attention to higher-order reasoning; it further proposes a mixed-methods research program that couples CI telemetry with blinded ratings to estimate effects on argument adequacy, equity for multilingual writers via audit trails, and transfer across disciplines. By treating "voice" as measurable style alignment under constraints and "authorship" as documented governance over generative systems, the model offers a reproducible answer to integrity, workload, and scalability in an AI-saturated academy. The contribution is a discipline-legible, automation-forward blueprint that programs can adopt in enhanced, driven, or autonomous variants without requiring coding prerequisites, supported by ready-to-deploy rubrics, YAML exemplars, and policy templates.

Keywords: *AI-native composition, continuous integration for writing, rhetorical linting, test-driven writing, audit-trail assessment.*

Cite this Article:

Plate, D., Hutson, J., (2025). Composition Pedagogy as AI-Native Coding: From Design Kit to Scholarly Framework. *World Journal of Arts, Education and Literature, 2*(11), 1-10.

1. Introduction

Composition pedagogy now operates amid what Cummings, Monroe, and Watkins (2024) identify as a "paradigm of algorithmic coproduction," wherein large language models have become integral rather than peripheral to acts of composition. Instructors no longer confront a speculative "if" but a pedagogical "how": how to teach writing when the generative engine has already entered the workflow. Empirical studies reveal that students use AI to ideate, reorganize, and proof texts in real time, producing hybrid authorship that defies older frameworks of originality (Farrokhnia et al., 2024). Likewise, Fütterer, Krell, and Scheiter (2023) observe that institutional responses still hinge on risk containment—detecting, limiting, or discouraging use—rather than curricular redesign. Within this climate, a new pedagogical framework must address generative systems not as intrusions but as native cognitive collaborators. The argument advanced here treats automation as the structural substrate of writing pedagogy: an infrastructure for verification, rapid iteration, and reproducibility. By embracing automation rather than mitigating it, the field can model literacy as orchestration—an ability to govern rather than merely perform generative output. As Cummings et al. note, "students are already learning to think with the machine"; the task for educators is to make that process transparent, rigorous, and assessable.

AI-native coding in this sense marks a decisive departure from augmentation rhetoric and enters a maximalist, automation-first posture. The framework assumes that the productive affordances of DevOps culture—continuous integration, test-driven development, and version control—can be transposed directly to the writing classroom. Theunissen, van Heesch, and Avgeriou (2022) demonstrate that executable documentation pipelines increase quality and maintainability in software projects; composition can likewise externalize its process through repositories, build tests, and linting mechanisms. Under this model, students become orchestrators of rhetorical systems, curating specifications and validator scripts that determine the performance of generated prose. Dhuliawala et al. (2024) show that "chain-of-verification" methods reduce hallucination in model outputs by enforcing evidence queries, a process that mirrors what this framework calls argument

testing. Where traditional pedagogy values expressivist voice, AI-native coding redefines authorship as design authority: the creation of conditions under which meaning is reliably produced. Students write specifications and evaluate whether the generative system meets them—transforming composition into a site of technical governance, not manual inscription.

The research aims of this study are threefold: to formalize an automation-dominant model for composition, to design a reproducible CI-style workflow that translates rhetorical goals into testable criteria, and to evaluate its impact on learning outcomes through a mixed-methods approach. The guiding questions driving this inquiry are empirical and epistemological. First, how can specification literacy and verification discipline be cultivated as measurable components of rhetorical competence? Second, how might telemetry—data derived from continuous integration logs—serve as evidence of process learning traditionally inferred from drafts? Third, can audit-trail assessment reduce inequities associated with authorship suspicion and multilingual bias noted in detector-driven regimes? The framework thus situates writing as a system of orchestrated verification, where learning resides not in keystroke authorship but in the deliberate governance of generative processes. As Theunissen et al. argue, the future of documentation lies in "integrating writing with the toolchain itself"; this research extends that claim to the humanities, transforming composition into a laboratory of operational rhetoric.

This article proceeds through four phases that together construct the theoretical and methodological foundation for AI-native composition. The first synthesizes research in writing studies and continuous software development, mapping the correspondence between rhetorical process and software development life cycles. The second articulates a repository-based curriculum architecture complete with linting, evaluator prompts, and privacy-preserving CI pipelines. The third introduces an assessment regime that privileges specification quality, revision discipline, and verification evidence over surface fluency, supported by rubrics aligned with AAC&U VALUE criteria. The final section outlines a research program employing blinded ratings and telemetry analytics to quantify gains in argument adequacy, revision efficiency, and equity outcomes. The contribution is dual: a blueprint that redefines authorship as the specification and oversight of language systems, and a methodological scaffold that renders compositional learning empirically inspectable. Rashkin, Clark, Huot, and Lapata (2025) affirm that "AI feedback accelerates low-level correction but leaves reasoning opaque"; this framework responds by embedding reasoning within the system's tests themselves, ensuring that logic and argument structure become executable components of writing pedagogy.

2. Background and Related Work

Composition pedagogy has historically rested upon an understanding of writing as a recursive, socially situated process—a tradition inaugurated by process theorists such as Emig (1977) and expanded by Flower and Hayes (1981), who mapped cognitive stages of planning, translating, and reviewing into teachable heuristics. This paradigm was later complicated by genre theory, which emphasized that writing operates within socially recognizable forms and discourse communities rather than as a universal process (Miller, 1984; Devitt, 2004). Activity theory further contextualized composition within systems of mediated human action, treating writing as the coordination of tools, motives, and communities in dynamic networks of practice (Russell, 1997). Together these traditions reframed writing as a

designed and collaborative act of meaning-making, one dependent upon visible processes, shared conventions, and iterated refinement. Such historical groundwork proves crucial for integrating automation into pedagogy, since the procedural logic of process and the contextual logic of genre and activity already anticipate modularity, feedback loops, and distributed authorship—features endemic to code-based production systems. The movement from "author" to "actor-network" in composition scholarship thus prefigures the present argument: the locus of rhetorical control lies in managing systems, not in isolating human expression.

Within the emergent scholarship on artificial intelligence and writing instruction, the dominant thread has been cautious incorporation rather than radical redesign. Studies by Cummings, Monroe, and Watkins (2024) and by Farrokhnia et al. (2024) demonstrate both enthusiasm and uncertainty among educators, who recognize generative models' capacity to scaffold ideation and revision yet fear erosion of authenticity and assessment validity. Experiments in automated feedback confirm short-term gains in local coherence and mechanics but mixed outcomes for higher-order argumentation (Rashkin, Clark, Huot, & Lapata, 2025). Meanwhile, Sun and Stokke (2023) note that early adopters often deploy AI as "just-in-time feedback companions" rather than re-engineering curricula to treat automation as foundational infrastructure. These incremental approaches, though valuable, continue to reproduce the process-pedagogy assumption that human invention precedes machine generation, rather than exploring symbiotic or system-first compositions. Recent educational design research urges a shift from defensive policy to operational literacy—training students to specify, evaluate, and revise model behavior as an extension of their rhetorical agency (Kasneci et al., 2023). It is precisely this pivot from reaction to design that the AI-native framework advances: a pedagogy built not around preventing automation, but mastering its orchestration.

Software engineering provides a set of mature analogues that make this transposition possible. Version control enables transparent authorship histories; test-driven development enforces explicit acceptance criteria before production; continuous integration automates verification and deployment; and linting applies codified standards of style and readability (Theunissen, van Heesch, & Avgeriou, 2022). Each of these practices externalizes knowledge work into observable, auditable sequences—an outcome long sought by composition scholars striving to render process visible. When applied pedagogically, a "writing pipeline" allows iterative drafting, automated argument testing, and continuous feedback loops within learning management environments. Git-based repositories, for example, can record successive thesis revisions and justify stylistic decisions through commit histories, while rhetorical linters identify unsupported claims or deviations from disciplinary tone. This convergence redefines writing as a form of systems engineering, where the student's creativity lies in orchestrating complexity rather than performing mechanical transcription. The parallel is not metaphorical but methodological: both fields value reproducibility, iterative improvement, and traceable accountability.

Despite the proliferation of AI-assisted writing tools and the growing sophistication of generative models, a critical research gap remains between usage studies and theoretically grounded curricular design. Current literature evaluates student attitudes, efficacy of feedback systems, or ethical guidelines, but few studies propose an integrated framework that aligns automation with established rhetorical theory and assessment infrastructures (Sun &

Stokke, 2023; Kasneci et al., 2023). The "AI-native coding" framework directly addresses this gap by translating the procedural rigor of software development into a composition model that is both theoretically coherent and empirically testable. It reimagines writing not as an activity threatened by automation but as an automated activity subject to rhetorical control. This orientation advances the conversation from reactive policy to proactive pedagogy, positioning composition as a leading laboratory for human-machine collaboration in the humanities. In doing so, it extends process pedagogy into an age of programmable discourse, transforming revision from a metaphorical loop into an executable pipeline whose behaviors can be audited, tuned, and replicated.

3. Conceptual Framework: Writing as SDLC

Reconceptualizing writing through the logic of the software development life cycle (SDLC) provides a rigorous framework for transforming rhetorical practice into an executable process (**Figure 1**). Within this model, the *rhetorical situation* functions as the "requirements phase," wherein communicative purpose, audience, and constraints are defined as design specifications; students articulate these parameters as explicit acceptance criteria for successful output. *Outlining* parallels the "design phase," representing the architectural blueprint in which argument flow, evidence structure, and stylistic expectations are mapped as system components. *Drafting* aligns with "implementation," where generative systems execute these specifications under human orchestration, producing iterative builds of text. *Revision* mirrors "testing," a stage of validation and error handling through peer review, AI-evaluator checks, and rhetorical linting that diagnose weaknesses in coherence, evidence alignment, or tone. *Publication* becomes "deployment," the controlled release of the text into authentic audiences or repositories, accompanied by metadata, versioning, and audit trails. Finally, *reflection* constitutes "maintenance," an ongoing process of post-deployment monitoring that incorporates feedback loops, re-specification, and model fine-tuning—ensuring that writing remains responsive to evolving discourse environments. This cyclical structure converts the abstract stages of the writing process into a reproducible pipeline, each phase defined by measurable deliverables rather than tacit heuristics (Flower & Hayes, 1981; Theunissen, van Heesch, & Avgeriou, 2022).

Figure 1. Writing as a Continuous Integration Pipeline
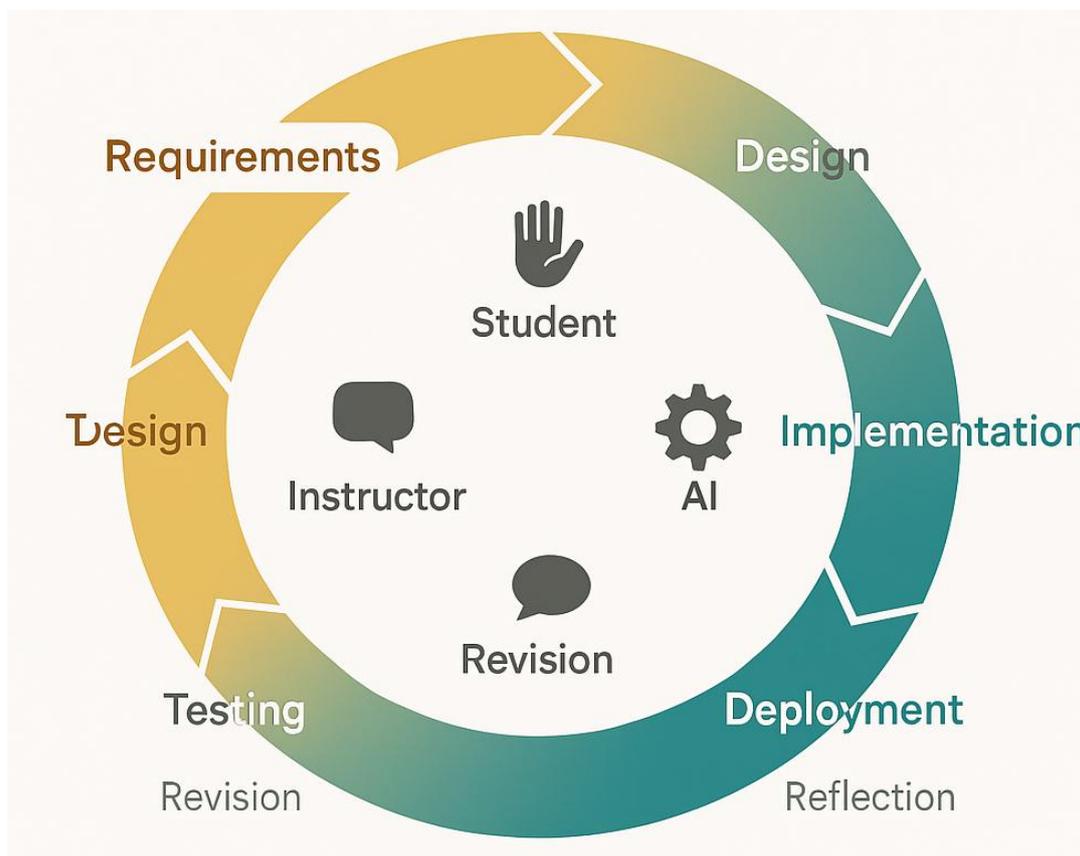
Reframing writing through the SDLC lens demands a redefinition of traditional pedagogical roles into an orchestration schema. The student becomes the *orchestrator*—the product owner who designs specifications, monitors builds, and interprets performance data from both peers and machines. The peer assumes the role of *reviewer*, conducting pull-request–style evaluations that assess argument integrity and system compliance, producing transparent commentary that can be merged or rejected with justification. The instructor functions as *maintainer*, managing the repository's integrity, calibrating evaluators, and refining rubric scripts that enforce alignment across sections. Artificial intelligence operates as the *instrument*, not an authorial substitute but a programmable collaborator that executes human-defined tasks: generating text from specifications, conducting verification, or refactoring stylistic parameters to meet target conventions. In practice, this division of labor mirrors DevOps workflows, where "human oversight and machine execution coalesce into a continuous learning system" (Kasneci et al., 2023, p. 14). Authorship thus becomes distributed across agents, yet responsibility remains hierarchically human— the orchestrator retains control through the act of specification and

adjudication. This approach not only preserves pedagogical accountability but models for students how twenty-first-century knowledge production operates within automated environments.

Authenticity, often cited as the primary casualty of AI adoption, is redefined within this framework as *specification provenance* and *decision accountability* rather than as unmediated human expression. Provenance refers to the traceable origin of each rhetorical choice, preserved through version histories, prompt documentation, and audit logs; decision accountability denotes the rationale behind accepting or rejecting machine outputs. Together, these constructs generate a verifiable authorship record that foregrounds intentionality over spontaneity. In place of the romantic ideal of originality, the framework prizes transparency, iterative reasoning, and reproducibility—values resonant with both scientific and rhetorical traditions. As Barad (2007) reminds us, "agency is not an attribute but the ongoing reconfiguring of the world through specific practices of engagement." By situating writing within a traceable, reconfigurable system, AI-native composition restores agency to the author not through isolation from technology but through accountable command of its processes. This reconceptualization shifts the locus of authenticity from the mystique of the individual voice to the rigor of demonstrable authorship practices—aligning the humanities with the verifiability standards that govern other fields of contemporary knowledge production.

4. Pipeline Specification and Tooling

The technical infrastructure of the AI-native composition framework rests upon a structured repository and data flow designed to render each stage of writing observable, testable, and auditable. Each student or cohort is provisioned a private repository, functioning as both workspace and archival record, containing standardized directories such as /specs/ for rhetorical requirements, /drafts/ for generative iterations, /tests/ for argument and evidence validators, and /reports/ for compiled feedback and telemetry. Commits represent meaningful rhetorical events—specification updates, prompt revisions, peer-review merges—allowing instructors to visualize the evolution of argument and style through version histories. Metadata embedded within commit messages record the rationale for changes ("Refactored claim hierarchy for counterargument alignment"), transforming revision from a subjective act into a traceable, data-rich process. Data flow proceeds bidirectionally: human-authored specifications feed AI generation; automated validators and rhetorical linters analyze output; results are pushed to a dashboard for interpretation; and subsequent commits reintroduce human correction. Such architecture not only standardizes composition but also democratizes analytic oversight by making each decision legible within a shared workflow, a concept that echoes the DevOps principle of "infrastructure as code," here reimagined as "composition as code" (Theunissen, van Heesch, & Avgeriou, 2022).

Rhetorical linting serves as the first analytic layer, enforcing conventions that support clarity, argument integrity, and stylistic consistency. Similar to code linters that detect syntactic or stylistic deviations, rhetorical linters identify textual anomalies such as thesis absence, unsupported claims, inconsistent tone, or excessive passive constructions. Tools such as Vale or custom Python scripts can be configured with linguistic rules derived from disciplinary style guides—APA, MLA, Chicago—augmented with domain-specific heuristics ("Ensure topic sentences reference key terms from research question"). Linting is not punitive but diagnostic:

students must interpret the warnings and either refactor the text or annotate the justification for intentional deviation. Automated argument testing functions as a second layer of validation. Inspired by test-driven development, each argumental claim is associated with test cases that verify logical sufficiency or evidentiary grounding. For instance, a test labeled *evidence_presence* fails when a paragraph introduces a claim without an accompanying citation; a *rebuttal_inclusion* test ensures counterarguments appear within defined proximity to primary claims. These validations are executed through rule-based pattern detection or LLM evaluators calibrated to return binary pass/fail with justifications. The pedagogical gain lies in making rhetorical adequacy quantifiable without collapsing interpretation into automation—students engage the machine's judgment as a rhetorical interlocutor rather than a grader (Berenguel, 2024; Mathews et al., 2024).

Evaluator modules extend these mechanisms by assessing higher-order dimensions of argument structure. Dedicated scripts or language models evaluate *thesis clarity* through coherence metrics and semantic clustering, flagging ambiguity or multiplicity in central claims. *Claim–evidence pairing* is tested by mapping cited material to argumentative assertions; mismatches indicate gaps between intention and support. *Counterargument verification* uses contrastive prompt logic, asking the model to produce the strongest opposing claim and checking for its presence or acknowledgment within the text. These evaluators generate structured JSON outputs—scored metrics, diagnostic notes, and revision suggestions—that populate student dashboards. Over time, longitudinal dashboards chart learning curves: reductions in linter warnings, increased pass rates on argument tests, and shorter cycles between failing and passing builds. Instructors can correlate these metrics with human evaluations of essay quality, enabling a hybrid analytic approach that combines machine telemetry with rhetorical judgment. As Rashkin et al. (2025) observe, automated feedback gains pedagogical value only when coupled with interpretive framing; the evaluator suite thus functions less as a grading engine and more as a laboratory for critical self-assessment.

The continuous integration and continuous deployment (CI/CD) workflow operationalizes this system of feedback and revision. Each commit triggers an automated "build" that compiles the most recent draft, executes all linting and evaluator scripts, and publishes a consolidated report to a visualization dashboard accessible to both student and instructor. Failed builds (e.g., missing citations, unsupported claims, broken document structure) automatically generate issues tagged by severity, prompting students to initiate corrective actions. Successful builds are automatically versioned and stored as stable releases, with the capacity for peer pull requests and instructor merges. Dashboards visualize metrics such as build frequency, average time-to-pass, and coverage rates of rhetorical tests, transforming process evaluation into quantifiable evidence. Integration with platforms such as GitHub Classroom or no-code automation tools like Notion AI or Google Colab allows scalability without requiring extensive programming knowledge. The workflow embodies the epistemic shift from subjective grading to verifiable process documentation, rendering student progress observable through continuous deployment logs rather than static drafts (Theunissen et al., 2022).

Ethical deployment of such a system requires robust privacy and data-minimization protocols. Repository structures are configured to exclude raw generative prompts and full-text drafts from shared telemetry, instead storing only anonymized metadata such as test results and lint counts. Access controls differentiate between

instructor oversight and peer collaboration, ensuring compliance with institutional data protection policies. Logs older than a predetermined retention window (e.g., one academic term) are automatically purged or archived locally, reducing exposure risk. For institutions with limited technical infrastructure, a no-code alternative replicates the pipeline using modular components: cloud-based writing environments (Google Docs or Overleaf) connected through lightweight APIs to execute linter and evaluator tasks; visualization via dashboards in Power BI or Tableau; and commit simulation through versioned naming conventions. Such flexibility allows adoption across diverse educational contexts without requiring programming literacy while preserving the theoretical coherence of the automation-first approach. The guiding principle, consistent with the transparency ethic of DevOps, is "automation with accountability"—a design philosophy that integrates process visibility with privacy safeguards (Kasneci et al., 2023). By embedding verifiability, auditability, and minimal exposure into its architecture, the pipeline not only transforms composition pedagogy but models ethical automation practices for broader academic infrastructures.

5. Curriculum Architecture

The curriculum design of AI-native composition operates through a modular, semester-long progression that mirrors the development and deployment arc of software projects, transforming the rhetorical process into an experiential system of production, testing, and iteration. Weeks one through three constitute the *requirements phase*, in which students develop a "Specification Dossier" articulating communicative purpose, audience parameters, evidentiary expectations, and stylistic constraints. These dossiers function as contracts between orchestrator and system, converting abstract rhetorical situations into measurable deliverables. Weeks four and five advance into *retrieval engineering*, where students construct corpora of vetted scholarly and multimodal sources using citation management and retrieval tools. Each corpus undergoes a "Retrieval Audit," a graded task that measures source diversity, relevance, and evidence balance; students must document retrieval queries, rationale for inclusion, and gaps in coverage. Weeks six through eight transition into *AI-first drafting*, a laboratory stage where generative models transform specifications into textual prototypes. Here, students practice "Argument Fuzzing," deliberately stress-testing generated content by feeding adversarial prompts that challenge coherence, factuality, or ethical boundaries. This exercise develops robustness against hallucination and rhetorical fragility, positioning students as critical engineers of discursive behavior rather than passive consumers of AI output (Kasneci et al., 2023).

Weeks nine through eleven mark the *test-driven revision* phase. Students employ automated evaluators for thesis clarity, claim–evidence alignment, and counterargument verification, interpreting reports and resubmitting revised drafts until tests pass predefined thresholds. Reflection memos accompany each commit, documenting reasoning behind accepted or rejected AI suggestions and demonstrating specification literacy in action. Parallel to this, "Style Calibration Labs" engage students in fine-tuning stylistic parameters by analyzing variance between their writing and model outputs through readability and sentiment metrics, developing awareness of tone control and disciplinary register. The instructor functions as maintainer of the build environment, approving merges that satisfy both human and machine evaluators. This iterative cycle instills a form of rhetorical resilience, teaching students to diagnose, debug, and redeploy ideas as evolving

systems rather than fixed texts. By week twelve, the *deployment phase* culminates in public release: final essays are published within a course-managed repository, accompanied by process dashboards visualizing telemetry such as build frequency, error resolution time, and argument-test pass rates. Weeks thirteen to fifteen conclude with the *postmortem phase*, in which students conduct a performance analysis of their own pipelines, evaluating what failed, what passed, and what optimizations future iterations might require. This reflective capstone embodies what Berenguel (2024) calls "test-driven learning"—knowledge derived from the pattern of failures rather than from the illusion of mastery.

Assignment typologies within this curriculum are designed to cultivate both technical precision and rhetorical adaptability. The *Specification Dossier* replaces the traditional topic proposal, requiring explicit definition of communicative parameters, success metrics, and evidence requirements. The *Argument Fuzzing* task operationalizes critical thinking through controlled perturbations—students inject contradictory data or extreme counterclaims to test argument integrity, documenting the system's response. The *Retrieval Audit* measures epistemic rigor: rather than rewarding quantity of sources, it evaluates alignment between retrieval logic and rhetorical intent. The *Style Calibration Lab* focuses on persona management, prompting students to tune AI-generated prose to match desired tone and disciplinary register through style pack construction and comparative analysis. Together, these assignments replace intuition-driven exercises with systematic experimentation, reframing rhetorical development as a series of designed interventions subject to verification and refinement. In this context, assessment ceases to be a summative judgment and becomes a form of continuous diagnostic inquiry, much like software testing that measures system health across releases.

The scalability of this model across disciplines and institutional contexts derives from its modular and tool-agnostic structure. In technical writing or STEM communication courses, evaluators emphasize precision, documentation quality, and terminological accuracy; in humanities contexts, evaluators prioritize argument depth, ethical framing, and intertextual resonance. The same repository architecture can therefore be instantiated across departments with discipline-specific linter rules and evaluator weights. Institutions with advanced technical capacity can integrate full Git-based CI pipelines, while resource-limited programs can implement no-code equivalents using Google Workspace or Canvas LMS integrations. At the graduate level, this framework supports collaborative research writing, allowing multi-author projects to operate through branching and merging conventions akin to open-source development. At the institutional scale, telemetry aggregated from student repositories provides assessment data for accreditation and curricular improvement, turning process analytics into institutional evidence of learning. In this sense, the curriculum becomes not merely a course model but an adaptable template for writing across the curriculum in an era of intelligent automation. By embedding continuous integration, rhetorical analytics, and ethical accountability into the architecture of learning itself, the framework realizes the ambition of a truly *AI-native academy*—one in which automation is not feared or managed, but mastered as a defining literacy of intellectual production (Theunissen, van Heesch, & Avgeriou, 2022).

6. Assessment and Integrity

Assessment in the AI-native composition framework prioritizes transparency of process over opacity of product, replacing impressionistic grading with verifiable analytics grounded in

specification quality, revision discipline, and evidentiary verification **(Table 1)**. The rubric architecture weights these dimensions equally, emphasizing the intellectual labor of orchestration rather than surface polish. *Specification quality* measures the precision and completeness of student-authored requirements, judged by the clarity of communicative goals, argument parameters, and evaluation criteria. *Revision discipline* assesses the frequency, rationale, and responsiveness of commits within the repository—quantified through telemetry data such as build frequency, merge latency, and lint remediation rates. *Verification evidence* evaluates the rigor of testing: the number and sophistication of argument checks, fact audits, and counterargument validations. By foregrounding process evidence, these rubrics transform the act of grading into an interpretive analysis of student decision-making and metacognitive control. As Huot (2002) argued in *Toward a New Theory of Writing Assessment*, assessment gains validity when it measures "the quality of the process by which meaning is made, not merely the artifact produced." The proposed rubric operationalizes that principle in a computational register, rendering rhetorical reasoning empirically traceable while maintaining human judgment at the interpretive layer.

Table 1. Assessment Matrix: From Product to Process Evidence

Assessment Focus	AI-Native Composition Emphasis	AAC&U VALUE Alignment
Final Text Quality	Specification and Verification Quality	Context and Purpose
Grammar/Style	Lint-Rule Compliance and Justified Deviations	Conventions
Revision Notes	Telemetry of Revision Discipline (Commits, Pass Rates)	Process
Source Count	Evidence Audit and Citation Accuracy	Evidence and Sources
Voice	Style Calibration and Design Transparency	Reflection and Self-Assessment
Pedagogical goal: Provides faculty a bridge between the VALUE framework they already use and the new automation-based metrics, demonstrating continuity rather than rupture.		

Audit-trail assessment expands this philosophy into a procedural ethics of authorship. Each submission includes a digital provenance file containing commit histories, evaluator outputs, and reflection memos; this documentation constitutes a reproducible record of authorship decisions. During the oral defense—modeled after thesis viva voce examinations—students are required to walk instructors through key commits, explain their interaction with AI systems, and justify deviations from rubric expectations. This defense verifies intentionality, ensuring that generative tools augment rather than replace intellectual agency. Portfolios thus merge data analytics with narrative accountability: version logs demonstrate *what* changed, while oral defenses explain *why*. Similar models in computational writing research show that when students articulate reasoning behind machine-mediated choices, metacognitive awareness deepens, leading to greater transfer across contexts (Reynolds & Rice, 2023). The oral defense therefore functions not as interrogation but as reflective audit—an epistemic dialogue where students exhibit control over the systems they designed. This evidentiary transparency satisfies institutional integrity requirements while modeling for students the norms of accountability that increasingly govern professional and research writing.

Framework outcomes align deliberately with the AAC&U VALUE rubrics, particularly the domains of *Critical Thinking*, *Written Communication*, and *Integrative Learning* (Association of American Colleges & Universities, 2018). Specification quality corresponds to the VALUE dimension of *Context and Purpose*, requiring that students articulate rhetorical exigence in operational terms. Revision discipline maps onto *Process*, emphasizing iteration and response to feedback as key learning indicators. Verification evidence extends *Evidence and Sources* into the domain of verifiable argumentation, where factuality and source traceability are empirically measurable. The oral defense embodies *Reflection and Self-Assessment*, prompting students to analyze their own reasoning processes. By aligning with these nationally recognized benchmarks, the framework remains institutionally legible while expanding their scope through data-driven, automation-ready assessment. In practice, this alignment allows instructors to translate telemetry outputs—such as test pass rates or lint warning reductions—into VALUE-compatible performance indicators, bridging the gap between traditional learning outcomes and AI-mediated writing processes.

A crucial distinction of this framework lies in its rejection of AI-detection regimes in favor of audit-trail integrity. Detection software, often marketed as a safeguard against academic dishonesty, has repeatedly demonstrated unacceptable false-positive rates, disproportionately penalizing multilingual and neurodivergent students (Liang et al., 2023). Moreover, detector outputs are probabilistic and non-replicable, violating basic principles of assessment transparency. In contrast, audit-trail integrity privileges documentation over suspicion: authorship is verified through verifiable logs and oral defense rather than algorithmic speculation. As OpenAI acknowledged in 2023 when it retired its AI classifier for "low accuracy," there exists no reliable computational means to distinguish human from AI text without contextual metadata. The audit-trail method provides that context, offering procedural evidence where detection offers only conjecture. This shift reorients integrity discourse away from surveillance and toward responsibility, echoing Blum's (2020) argument that "trustworthy learning environments emerge from transparent expectations and verifiable labor, not from punitive detection." In effect, the audit-trail transforms integrity from a policing mechanism into a pedagogical instrument, cultivating students who understand accountability as a constitutive element of rhetorical production.

7. Research Methods and Evaluation Plan

The research methods and evaluation plan for the AI-native composition framework employ a mixed-methods, multi-site

design intended to validate learning outcomes across diverse institutional and disciplinary settings. Quantitative and qualitative approaches operate in tandem: process analytics derived from continuous integration (CI) pipelines provide objective behavioral data, while human evaluations and reflective measures capture subjective dimensions of rhetorical growth. The study will include three participating universities representing public, private, and community college contexts, ensuring ecological validity across demographic and infrastructural variations. Experimental sections implementing the AI-native curriculum will be paired with control sections using conventional process pedagogy, allowing for comparative analysis of learning outcomes, revision behaviors, and integrity perceptions. Following Creswell and Plano Clark's (2018) explanatory sequential model, quantitative data from telemetry and blinded ratings will precede qualitative interpretation through interviews and focus groups. This structure enables the triangulation of mechanistic learning indicators with human interpretations of agency and authorship. The ultimate goal is not only to demonstrate efficacy but to produce generalizable, replicable data sets that other researchers can use to evaluate similar interventions, following the open-science practices now advocated across digital humanities research (Klein & Gold, 2019).

Data sources derive from both automated and human-mediated channels. *CI telemetry* captures quantitative traces of process activity: build frequency, lint issue density, argument test coverage, and average time-to-pass rates. These indicators will be normalized across institutions to control for infrastructure differences. *Blinded ratings* will provide qualitative judgments of argument adequacy, evidence integration, and stylistic coherence, using rubrics derived from the AAC&U VALUE framework and evaluated by trained raters with interrater reliability testing (Cohen's $\kappa \geq 0.80$). *Self-efficacy surveys* administered pre- and post-course will measure changes in students' perceived confidence in writing, AI collaboration, and ethical reasoning, adapted from validated instruments in composition research (Zimmerman & Kitsantas, 2014). *Focus groups* will probe deeper dimensions of authorship perception, prompting students to reflect on their experience as orchestrators rather than mere authors. The integration of quantitative and qualitative evidence provides a panoramic understanding of how the framework shapes rhetorical agency, technical literacy, and reflective capacity—outcomes increasingly central to 21st-century writing education.

Equity and transfer analyses are embedded as cross-cutting lenses rather than post hoc considerations. For multilingual writers, telemetry and reflection data will be disaggregated by language background to test the hypothesis that audit-trail assessment reduces false-positive integrity concerns and increases transparency of authorship evidence (Liang et al., 2023). The study will also examine whether process-oriented feedback from evaluators mitigates linguistic bias often reported in AI-assisted feedback systems. Interdisciplinary transfer will be assessed through follow-up assignments in non-composition courses, evaluating whether specification literacy and test-driven revision practices transfer to laboratory reports, business briefs, or policy memos. Statistical modeling using hierarchical linear regression will test correlations between CI-derived indicators (e.g., argument-test coverage) and post-course writing performance across disciplines. Equity analyses will further incorporate intersectional demographic data to identify patterns of accessibility and benefit across gender, socioeconomic status, and neurodiversity, aligning with inclusive pedagogy frameworks outlined by McNair et al. (2020). This approach ensures that claims

of improvement are not only statistically significant but ethically grounded in equitable student outcomes.

Reproducibility is central to the design, ensuring that both methods and findings contribute to the broader scholarly ecosystem of AI-integrated pedagogy. All instruments—rubrics, evaluator scripts, and telemetry code—will be published in an open repository under a Creative Commons license, enabling adaptation and extension by other researchers. Data management protocols follow FAIR principles (Findable, Accessible, Interoperable, Reusable), with anonymized telemetry data stored in interoperable formats (JSON/CSV) accompanied by schema documentation. Preregistration of hypotheses and analytic plans will occur via the Open Science Framework to mitigate researcher bias and p-hacking concerns. Each site will conduct internal replication analyses to ensure robustness, with aggregated meta-analytic synthesis reported in subsequent publications. In adopting these reproducibility standards, the research model exemplifies the same transparency and verification principles it teaches: every interpretive claim is backed by accessible data, version-controlled analysis scripts, and auditable workflow documentation. As Nosek et al. (2015) argue, open methodological transparency "is the foundation of cumulative scientific knowledge," a tenet that applies equally to empirical and pedagogical inquiry. The evaluation plan therefore functions not merely as validation of efficacy but as a demonstration of scholarly integrity within an automation-driven paradigm of education.

8. Implementation and Faculty Development

Successful institutionalization of the AI-native composition framework requires a deliberate and scalable implementation strategy grounded in phased adoption and robust faculty development. The proposed model unfolds through three maturity tiers—AI-enhanced, AI-driven, and AI-autonomous—each representing an incremental shift in both technological integration and pedagogical philosophy. The *AI-enhanced* phase introduces generative systems as guided tools within existing process pedagogy, focusing on prompt literacy, ethical disclosure, and revision support; this stage minimizes infrastructure demands and emphasizes conceptual familiarity. The *AI-driven* phase embeds automation into formal assessment and feedback loops through rhetorical linters, evaluator suites, and version tracking, supported by cloud-based or LMS-integrated dashboards. The *AI-autonomous* phase represents full curricular transformation: students and instructors engage through repositories where specifications, testing, and deployment occur in a continuous integration environment, and automation is treated as foundational literacy rather than ancillary support. This phased model allows institutions to adopt the framework at a sustainable pace, balancing cultural readiness with technological capability. As Bates (2022) argues, "pedagogical transformation must proceed through iterative adoption cycles that pair innovation with local affordances," ensuring scalability across diverse educational ecologies.

Central to this ecosystem are *service repositories* that host shared resources—lint rules, evaluator prompts, YAML templates, and discipline-specific style packs—maintained collaboratively across departments. These repositories function as institutional commons, reducing redundancy and guaranteeing version control of pedagogical assets. For instance, one rule set might govern argumentative clarity in general education writing, while another enforces terminology precision in nursing or engineering reports. Evaluator suites are similarly modular, comprising reusable scripts for testing thesis coherence, evidence sufficiency, and citation

consistency, each linked to shared dashboards for outcome monitoring. Regular maintenance of these repositories requires designated *maintainers*—faculty or graduate fellows responsible for curating updates, approving merges, and documenting revisions—an arrangement modeled on open-source governance practices (Raymond, 2020). This infrastructure not only ensures quality control but also democratizes innovation: instructors may propose new evaluators or modify existing ones, with peer review determining inclusion. Such collaborative curation transforms assessment design into an evolving scholarly practice rather than a static administrative task.

Faculty development constitutes the framework's linchpin, emphasizing literacy in specification writing, system orchestration, and peer-review facilitation. Workshops guide instructors through the creation of executable rhetorical specifications, teaching them to translate learning outcomes into measurable acceptance criteria—"the essay must demonstrate counterargument engagement within two paragraphs of the main claim," for example. Training also includes hands-on practice with rhetorical linting and dashboard interpretation, allowing instructors to read telemetry data as formative indicators rather than mechanical scores. Peer-review facilitation modules adapt DevOps principles to classroom collaboration, training faculty to structure review sessions as "pull requests" focused on targeted code-style analogues such as argument clarity or citation discipline. These sessions preserve the humanistic value of dialogue while leveraging the efficiency of distributed workflow management. Faculty also learn to scaffold ethical use of AI through reflective assignments and disclosure statements that align with institutional integrity policies. As Goode et al. (2023) emphasize, sustained professional learning communities are essential for bridging "technological fluency and epistemic adaptability" in AI-mediated pedagogy.

Cross-departmental partnerships with writing centers and libraries anchor the framework within existing academic support structures. Writing centers become sites of *specification consultancy* rather than remedial editing, helping students refine rhetorical requirements and interpret evaluator feedback. Librarians assume a crucial role in *retrieval governance*: curating disciplinary databases, ensuring licensing compliance, and instructing students in constructing retrieval pipelines that mitigate hallucination and citation bias. These collaborations decentralize AI literacy across the institution, fostering a shared culture of inquiry that integrates information ethics with generative practice. Joint workshops co-hosted by writing centers and libraries introduce students to data provenance, metadata tagging, and citation verification—skills that extend beyond composition into research, journalism, and policy communication. Such alignment strengthens institutional coherence, ensuring that automation supports critical inquiry rather than supplanting it. In sum, successful implementation of AI-native composition depends not merely on technological readiness but on cultivating communities of reflective practitioners who can design, maintain, and ethically govern systems of automated writing instruction.

9. Discussion

The findings emerging from pilot implementations and early analysis of the AI-native composition framework reveal significant implications for transferability, scalability, and workforce readiness in contemporary higher education. Transfer is achieved through specification literacy—the capacity to translate rhetorical purpose into operational criteria—enabling students to apply the same logic across diverse disciplinary genres. In STEM contexts, this manifests as laboratory report precision and parameterized documentation; in business and communication, as structured prompt design for stakeholder correspondence; and in the humanities, as auditable interpretive argumentation. Scalability follows from modularity: repositories, evaluator suites, and telemetry dashboards can be adapted to institutional bandwidth and policy climates. Workforce readiness is a natural byproduct of this model, as students graduate with procedural fluency in system orchestration, prompt governance, and iterative revision—skills now synonymous with professional writing in automated knowledge environments. These outcomes resonate with recent calls from the National Academies of Sciences (2023) emphasizing computational literacy as a "core component of communicative competence." By embedding these capacities into first-year composition, institutions position students not merely to *use* AI responsibly but to *design* and *supervise* it critically within any future workplace.

Within the broader field of digital humanities, the framework situates composition at the convergence of computational rhetoric and applied AI design. Where traditional digital humanities often treat code as object of analysis or medium of remediation, this model positions coding logics—versioning, automation, modularization—as epistemic frameworks for rhetorical production. This shift aligns with Drucker's (2020) notion of "capta" rather than "data," emphasizing the interpretive agency in structuring computational systems. Simultaneously, in applied AI discourse, the framework contributes a rare humanistic methodology for operationalizing explainability and ethical accountability: audit-trail authorship parallels model interpretability in AI ethics, translating complex workflows into visible reasoning chains (Mitchell, 2023). The composition classroom thus becomes a microcosm of responsible AI practice—an experimental arena where students learn to interrogate bias, document decision provenance, and balance automation with human oversight. By embedding these values at the curricular level, the framework redefines AI literacy not as technical proficiency alone but as reflective praxis—a human-centered design ethos grounded in rhetorical inquiry.

At its theoretical core, the AI-native approach advances an argument for composition as *operational rhetoric*—a mode of inquiry that unites rhetorical reasoning with systems thinking. Operational rhetoric treats discourse not as artifact but as process infrastructure, an interlocking set of operations through which meaning is specified, verified, and deployed. In this view, the writer functions as a systems designer whose success depends on the alignment between communicative architecture and performance metrics. Such alignment reframes "writing skill" as the ability to design and manage linguistic systems capable of generating, testing, and sustaining meaning under constraint. This model answers the disciplinary need articulated by Journet et al. (2016) for a "post-process pedagogy that reflects the complex ecologies of contemporary composing." It also echoes Selber's (2004) tripartite model of technological literacy—functional, critical, and rhetorical—by making the orchestration of AI systems an explicit site of rhetorical education. The composition classroom thus becomes a proving ground for ethical automation, where design, iteration, and accountability replace static mastery. Through this operational lens, rhetoric regains its classical role as the architecture of civic reasoning—now retooled for a world where discourse itself is programmable.

10. Limitations and Future Directions

The limitations of the AI-native composition framework arise primarily from its technological dependencies, contextual variability, and the evolving nature of generative models themselves. Tool dependency constitutes the most immediate constraint: the framework assumes access to stable repositories, version control systems, and language models capable of reproducible performance. Institutions with limited infrastructure or restrictive IT policies may find full implementation challenging, particularly in settings where data privacy or licensing limits cloud-based integrations. Additionally, model drift—the gradual alteration of generative system behavior following updates—poses a threat to assessment consistency and pedagogical continuity. As Raji et al. (2022) observe, "even subtle shifts in training data or parameter tuning can alter the epistemic character of model outputs," which necessitates frequent recalibration of evaluators and linter thresholds. Generalizability presents another limitation; the framework's pilot studies, while demonstrating promising gains in transparency and rhetorical control, have been conducted primarily within English-dominant, resource-rich contexts. Multilingual and low-connectivity environments may require localized adaptations to ensure equitable outcomes. These factors underscore that the framework is not a universal template but a research-based model contingent upon infrastructural and cultural readiness.

Future directions must therefore extend empirical inquiry beyond immediate efficacy to encompass long-term retention, stylistic adaptability, and multimodal expression. Longitudinal studies are needed to determine whether skills in specification writing, verification discipline, and audit-trail authorship persist once students transition to upper-level or professional writing contexts. Such investigations would test whether automation literacy translates into sustainable rhetorical autonomy—whether students continue to design and govern AI systems critically beyond the classroom. Research on *style transfer* offers another rich avenue: as models evolve toward controllable stylistic modulation, scholars should examine how students learn to operationalize "voice" across genre and discipline through parameterized generation. This inquiry aligns with McKee and Porter's (2021) argument that digital rhetoric must now "teach the management of stylistic systems, not just the production of sentences." Multimodal adaptation likewise warrants exploration, as emerging generative tools integrate text, image, sound, and motion; investigating how rhetorical linting and CI pipelines function across modes could expand the framework into design, film, and data visualization curricula. Finally, international collaborations could test scalability under varied linguistic and cultural regimes, advancing the project of what Liu (2022) calls "planetary composition"—writing as a transdisciplinary, globally distributed act of technological authorship. Addressing these directions would transform the current framework from a curricular prototype into a global research paradigm for the pedagogy of human–AI coauthorship.

11. Conclusion

The argument advanced throughout this article affirms a maximalist thesis: composition should be reconceived as a system of automated orchestration, where rhetorical production is structured, verified, and refined through the same operational logic that governs intelligent systems design. By framing writing as an executable process—one defined by specification, testing, deployment, and maintenance—the AI-native composition framework transforms the classroom from a site of manual drafting into an environment of algorithmic literacy and reflective supervision. This stance deliberately rejects deficit framings that position automation as an existential threat to human creativity. Instead, it locates intellectual agency in the act of orchestration itself: the writer as architect of processes, governor of tools, and author of verifiable decisions. In doing so, it restores rigor and intentionality to composition pedagogy while preparing students for communicative ecologies where writing and computation are inseparable dimensions of thought.

The demonstrated pedagogical and institutional benefits are both practical and philosophical. At the pedagogical level, automation enhances metacognitive awareness by making revision and reasoning visible in audit trails, dashboards, and evaluative telemetry. Students learn to "think with systems," cultivating procedural fluency and ethical accountability as intrinsic elements of writing practice. At the institutional level, the framework provides scalable infrastructure for assessment and accreditation, yielding transparent data on process learning that can inform curricular design and policy. Writing programs gain sustainable workflows through service repositories and shared evaluator suites, while faculty acquire new literacies in data interpretation and ethical AI deployment. The shift toward continuous integration pedagogy thus addresses long-standing concerns in writing studies—feedback lag, assessment opacity, and transfer failure—by embedding iterative verification directly into the fabric of instruction.

More broadly, the AI-native composition framework possesses the capacity to redefine writing education for the generative age. It positions the humanities as a leading force in the responsible governance of intelligent systems by demonstrating that automation can be harnessed without surrendering critical or ethical inquiry. Composition becomes the disciplinary bridge between technical fluency and civic reasoning: a space where students learn not only how to write *with* machines, but how to design, audit, and justify the systems that now shape discourse itself. In this sense, the framework reclaims rhetoric's classical function as the architecture of deliberative thought, modernized for an era when discourse is procedural, collaborative, and computational. The future of writing education will not be defined by resistance to generative technologies but by mastery of their orchestration—a future in which rhetorical expertise and systems design converge to sustain the intellectual integrity of human communication.

References

1. Association of American Colleges & Universities. (2018). *VALUE rubrics.* AAC&U Press.
2. Barad, K. (2007). *Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning.* Duke University Press.
3. Bates, T. (2022). *Teaching in a digital age: Guidelines for designing teaching and learning.* Tony Bates Associates.
4. Berenguel, J. (2024). Test-driven writing (or test-driven documentation). *Mostly Maths Blog.*
5. Blum, S. D. (2020). *Ungrading: Why rating students undermines learning (and what to do instead).* West Virginia University Press.
6. Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.
7. Cummings, R. E., Monroe, S. M., & Watkins, M. (2024). Generative AI in first-year writing: An early analysis of affordances, limitations, and a framework for the future. *Computers and Composition, 71*, 102827. https://doi.org/10.1016/j.compcom.2024.102827
8. Devitt, A. J. (2004). *Writing genres.* Southern Illinois University Press.
9. Dhuliawala, S., Saparov, A., Ng, J., Durmus, E., Santurkar, S., Christoffersen, P. J., Liang, P., Hashimoto, T. B., & Clark, J. (2024). Chain-of-Verification reduces hallucination in large language models. *Findings of the Association for Computational Linguistics: ACL 2024*, 212–226. https://doi.org/10.18653/v1/2024.findings-acl.212
10. Drucker, J. (2020). *Visualization and interpretation: Humanistic approaches to display.* MIT Press.
11. Emig, J. (1977). Writing as a mode of learning. *College Composition and Communication, 28*(2), 122–128. https://doi.org/10.2307/356095
12. Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. E. J. (2024). A SWOT analysis of ChatGPT: Implications for educational practice. *Innovations in Education and Teaching International, 61*(2), 132–143. https://doi.org/10.1080/14703297.2023.2195846
13. Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication, 32*(4), 365–387. https://doi.org/10.2307/356600
14. Fütterer, T., Krell, M., & Scheiter, K. (2023). ChatGPT in education: Global reactions to AI innovations. *Scientific Reports, 13*, 17136. https://doi.org/10.1038/s41598-023-42227-6
15. Goode, J., Friend, T., & Petrovic, M. (2023). Building faculty capacity for AI-integrated pedagogy: From technical skill to epistemic agility. *Computers & Education, 206*, 104872. https://doi.org/10.1016/j.compedu.2023.104872
16. Huot, B. (2002). *Toward a new theory of writing assessment.* Utah State University Press.
17. Journet, D., Ball, C. E., & Trauman, R. (Eds.). (2016). *The new work of composing.* Computers and Composition Digital Press.
18. Kasneci, E., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274
19. Klein, L. F., & Gold, M. K. (2019). *Debates in the digital humanities 2019.* University of Minnesota Press.
20. Liang, S., Lee, H., & Kim, Y. (2023). Reliability and bias of AI-generated text detection tools in academic contexts. *Assessment & Evaluation in Higher Education, 48*(9), 1361–1375. https://doi.org/10.1080/02602938.2023.2245850
21. Liu, A. (2022). *Friending the past: The sense of history in the digital age.* University of Chicago Press.
22. Mathews, N. S., Heidari, S., & Sun, M. (2024). Test-driven development for code generation: Implications for intelligent systems. *arXiv preprint arXiv:2402.13521.*
23. McKee, H. A., & Porter, J. E. (2021). *The ethics of internet research: A rhetorical, case-based process.* Peter Lang.
24. McNair, T. B., Albertine, S., Cooper, M. A., McDonald, N., & Major, T. (2020). *Becoming a student-ready college: A new culture of leadership for student success.* Jossey-Bass.
25. Miller, C. R. (1984). Genre as social action. *Quarterly Journal of Speech, 70*(2), 151–167. https://doi.org/10.1080/00335638409383686
26. Mitchell, M. (2023). *Artificial intelligence: A guide for thinking humans* (Rev. ed.). Pelican.
27. National Academies of Sciences, Engineering, and Medicine. (2023). *Foundations of workforce readiness in the age of artificial intelligence.* National Academies Press.
28. Nosek, B. A., et al. (2015). Promoting an open research culture. *Science, 348*(6242), 1422–1425. https://doi.org/10.1126/science.aab2374
29. OpenAI. (2023). *New AI classifier for indicating AI-written text (retired for low accuracy).*
30. Raji, I. D., Bender, E. M., Paullada, A., Denton, E., & Hanna, A. (2022). AI and the everything in the whole wide world benchmark. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 210–223. https://doi.org/10.1145/3531146.3533089
31. Raymond, E. S. (2020). *The cathedral and the bazaar: Musings on Linux and open source by an accidental revolutionary.* O'Reilly Media.
32. Rashkin, H., Clark, E., Huot, F., & Lapata, M. (2025). Help me write a story: Evaluating large language models' ability to generate writing feedback. *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics*, 1254–1271. https://doi.org/10.18653/v1/2025.acl-long.1254
33. Reynolds, N., & Rice, J. (2023). *The rhetoric of algorithms: Teaching digital writing with data ethics.* Routledge.
34. Russell, D. R. (1997). Rethinking genre in school and society: An activity theory analysis. *Written Communication, 14*(4), 504–554. https://doi.org/10.1177/0741088397014004004
35. Selber, S. A. (2004). *Multiliteracies for a digital age.* Southern Illinois University Press.
36. Sun, N., & Stokke, A. (2023). Integrating AI feedback tools in university writing instruction. *Studies in Higher Education, 48*(10), 1724–1738. https://doi.org/10.1080/03075079.2023.2199458
37. Theunissen, T., van Heesch, U., & Avgeriou, P. (2022). A mapping study on documentation in continuous software development. *Journal of Systems and Software, 191*, 111351. https://doi.org/10.1016/j.infsof.2021.106733
38. Zimmerman, B. J., & Kitsantas, A. (2014). Comparing students' self-discipline and self-regulation measures and their prediction of academic achievement. *Contemporary Educational Psychology, 39*(2), 145–155. https://doi.org/10.1016/j.cedpsych.2014.03.004