AI-Aesthetic TPACK: Redesigning Art Teacher Education in the Age of Generative AI

Lihui*

Ganesha University of Education
*Corresponding Author: 976621933@gmail.com

ABSTRACT

Generative text-to-image systems lower technical barriers to visual production but can also short-circuit learning if used end-to-end. This paper proposes AI-Aesthetic TPACK, a framework that aligns Human-in-the-Loop (HITL) and Human-out-of-the-Loop (HOTL) workflows to Leder's five stages of aesthetic processing (perceptual analysis, implicit memory integration, explicit classification, cognitive mastering, evaluation) and operationalizes it as a classroom toolkit: (a) a behavior-anchored competency rubric for teacher education; (b) a Process-Based Evidence Package (PBEP) template that makes creative processes visible and assessable; and (c) a course-level AI use & disclosure policy. Using a PRISMA-ScR–guided scoping synthesis, we distill design patterns that preserve stage-specific learning while leveraging AI for ideation and iteration. A 90-minute micro-trial with three experienced art instructors shows moderate inter-rater agreement when applying the rubric to a PBEP-documented student project (κoverall = 0.65; κdimension = 0.58–0.73). Qualitative debriefs reveal consistent pain points (e.g., ambiguity around licensing of texture assets) and lead to a shared casebook for future calibration. Contributions are: (1) a stage-aligned mapping of AI interventions with scaffold/shortcut conditions; (2) a ready-to-use toolkit (rubric + PBEP + policy) with evidence anchors; and (3) initial usability/consistency signals for short-format teaching. We conclude with boundary conditions, cultural considerations, and a data/materials availability statement to support replication and local adaptation.

Keywords: Aesthetic cognition; Generative AI; Teacher education; Technology integration; TPACK

INTRODUCTION

Generative AI systems—such as Midjourney, DALL·E, and Stable Diffusion—have become a transformative force across creative practice and education. In art and design classrooms, text-to-image models markedly lower technical barriers to visual production, enabling learners to generate complex, aesthetically refined imagery from simple textual prompts. In studio-based and design-oriented subjects, these systems are widely recognized as accelerators for ideation, rapid prototyping, and iterative exploration. At the same time, their rise poses new pedagogical challenges: educators must determine how to integrate AI tools in a way that enriches learning rather than undermining fundamental skills or creativity.

Within education research, technology integration is often framed by the Technological Pedagogical Content Knowledge (TPACK) framework, which formalizes how effective teaching arises from the dynamic interplay among content knowledge, pedagogy, and technology (Mishra & Koehler, 2006). TPACK provides a vocabulary for describing where AI can be positioned in a curriculum (for example, as a tool, a task modality, or an assessment resource) and how it interacts with teachers' pedagogical moves and disciplinary aims. From the perspective of cognitive and aesthetic psychology, Leder et al.'s five-stage model of aesthetic appreciation offers a stage-based account of how humans engage with art: perceptual analysis (analyzing formal elements), implicit memory integration (connecting to personal experiences), explicit classification (identifying style, genre, or context), cognitive mastering (resolving ambiguities or constructing meaning), and evaluation (forming an overall judgment) (Leder et al., 2004). Together, TPACK and the Leder model provide complementary lenses on AI's role in art education: TPACK clarifies the instructional locus of AI integration, while the Leder model clarifies the cognitive locus at which AI may scaffold or short-circuit learners' aesthetic reasoning. Despite these guiding frameworks, the intersection between instructional integration and aesthetic cognition remains inadequately specified for art and design education.

In particular, several critical issues are insufficiently theorized and operationalized in the current literature and practice:

Stage-specific mechanisms: We lack precise accounts of how generative AI interventions align with each stage of aesthetic cognition, and which learning operations (e.g. noticing, rule articulation, iterative refinement, judgment) are being genuinely scaffolded versus silently bypassed.

Scaffold versus shortcut conditions: It is unclear under what conditions AI functions as a productive scaffold rather than a counterproductive shortcut. The role of workflow design—especially HITL versus HOTL configurations—has not been systematically mapped to stage-specific learning affordances and risks. (Here, we define a Human-out-of-the-Loop (HOTL) workflow as one in which human input is minimal and the process relies on end-to-end AI generation of a final or near-final product, in contrast to Human-in-the-Loop (HITL) workflows that involve substantive human curation and editing.)

Assessment and evidence: Studio assessment in art education remains heavily artifact-centric; shared norms for process evidence (e.g. prompt logs, human edits and annotations, rationale for decisions, and AI-use disclosure) and stage-aligned evaluation criteria are scarce. Educators currently lack rubrics that reward how students think and iterate, not only what final artifact they submit.

Teacher competencies and policy: There is no behavior-anchored, domain-specific articulation of AI-Aesthetic TPACK competencies for teacher preparation. Likewise, we lack course-level policies for AI use and disclosure that translate academic integrity and copyright concerns into practical, teachable routines.

Design patterns for coursework: Practical lesson- and module-level patterns that operationalize HITL scaffolds (for example, "first human, then model" sequencing, two-step human–AI remix cycles, or "paired blind review + evidence check" activities) are only sporadically described in the literature and are rarely aligned explicitly to stages of aesthetic development.

These unresolved issues lead to a central problem: How should art teacher education be redesigned so that generative AI serves as a scaffold for stage-based aesthetic learning rather than short-circuiting it? Addressing this question is crucial for preparing teachers to harness AI in a pedagogically sound and ethically transparent way. In sum, educators need a coherent framework that (a) maps AI's scaffold-versus-shortcut potentials onto specific stages of aesthetic cognition; (b) specifies teacher competencies in an AI-Aesthetic TPACK rubric; and (c) operationalizes process-focused evidence collection and AI use/disclosure policies for classroom implementation. This need motivates the present study.

Purpose and Objectives: To respond to the above gaps, this study proposes a coherent framework and toolkit for AI-A-TPACK in art teacher education. The aim is twofold. First, we develop and justify a set of practical instruments – including a behavior-anchored competency rubric, a process-based evidence template, and a model AI use/disclosure policy – that together operationalize AI-A-TPACK for classroom use. Each component of this toolkit is grounded in the TPACK and aesthetic cognition frameworks, as we explain in later sections. Second, we conduct an initial exploration of the toolkit's usability and consistency in a short-duration teaching context. In particular, we pilot the rubric and evidence template in a brief workshop scenario to examine whether teachers can apply the rubric reliably (i.e. with consistent assessments) within the constraints of a single class session. By pursuing these two objectives, the study seeks to provide both a theoretically sound framework and early empirical insight into its practical viability for art and design teacher education.

THEORETICAL FRAMEWORK

TPACK in Art and Design Education

The Technological Pedagogical Content Knowledge framework (TPACK; Mishra & Koehler, 2006) posits that effective teaching with technology arises from an intersection of content knowledge (CK), pedagogical knowledge (PK), and technological knowledge (TK). In art and design education, this model helps conceptualize how emerging technologies like generative AI can be woven into pedagogy. For art teachers, AI-specific technological knowledge includes understanding the capabilities and limitations of generative models (e.g. knowing what text-to-image systems can and cannot easily do). AI-related pedagogical knowledge involves knowing how to integrate these tools into teaching strategies and studio routines (e.g. when to allow or encourage AI use during an art-making process, and how to scaffold that use). AI-related content knowledge entails understanding how AI relates to artistic content and creative processes (for instance, knowing how style transfer or image generation can illustrate art concepts or techniques). An AI-augmented extension of TPACK – which we refer to as AI-Aesthetic TPACK (AI-审美TPACK) or AI-A-TPACK – thus adapts the traditional framework to account for how AI tools mediate creative content and methods. It provides a structured lens to identify where in the curriculum AI acts (for example, as a tool for generating visual elements, as part of a pedagogical approach to critique and iteration, or as content in discussions about art and technology) and how it interacts with pedagogical intent and artistic learning goals. TPACK offers stability in terminology and design, ensuring that when we introduce AI, we consider not just the tool itself, but how it fits with pedagogy and content objectives. In summary, AI-A-TPACK emphasizes that meaningful integration of AI in art education requires aligning technical possibilities with sound pedagogical strategies and deep content understanding.

Leder's Model of Aesthetic Cognition

Complementing TPACK's educational lens, Leder et al.'s five-stage model of aesthetic appreciation (Leder et al., 2004) provides a cognitive–affective map of how individuals engage with art. The stages progress from basic perception to deeper evaluation, offering a sequence for the viewer's or creator's experience: (1) Perceptual analysis – noticing formal elements such as line, color, composition, and other basic features; (2) Implicit memory integration – making personal associations and drawing on prior experiences or knowledge in response to the work; (3) Explicit classification – identifying style, genre, or contextual information (for example, recognizing a Baroque painting style or a specific artist's hallmark techniques); (4) Cognitive mastering – resolving interpretive challenges or ambiguities, constructing meaning and understanding the work in a broader context; and (5) Evaluation – forming a judgment about the artwork's quality, success, or emotional impact. This model has been widely applied in art education research to align teaching strategies or critique methods with the cognitive processes students undergo when creating or viewing art. For instance, a teacher might scaffold the perceptual analysis stage by training students to carefully observe and verbally describe artworks before moving on to interpretation. Leder's framework thus informs what kind of thinking or appreciation skill is being targeted at each step of an art-making or art-viewing activity. In our study, these stages serve as a backbone for analyzing where AI might support or hinder learning – essentially linking the cognitive "locus" of AI's impact to specific pedagogical interventions.

Generative AI in Art Education: Current Insights

Recent research on AI in art and design education spans several strands, which together paint a picture of opportunities and challenges:

Implementation studies: many classroom implementation reports describe how text-to-image systems are introduced into studio tasks for ideation, prototyping, and iteration within art/design curricula. These studies often highlight AI's benefits in expanding the range of ideas students can explore and lowering the barrier for creating visual drafts (e.g., Wen & Wen, 2024). AI tools have been shown to help students generate rapid visual variations and explore creative options that they might not have attempted manually, thereby functioning as idea catalysts.

Ethical and policy analyses: Parallel research has examined issues of authorship, attribution, bias, and intellectual property raised by generative pipelines. For example, concerns are discussed regarding how students and educators should credit AI contributions and avoid plagiarism or misuse of copyrighted training data (Yusuf et al., 2024). This line of work emphasizes the need for clear guidelines and policies around AI usage in an educational context, to maintain academic integrity and address legal/moral questions (such as the permissibility of using AI-generated images in student artwork or the fairness of AI-derived content).

Perception and evaluation studies: Another strand investigates how viewers and judges respond to artwork that involves AI in the creation process, comparing perceptions of human-made, fully AI-generated, and hybrid human–AI works. For example, Agudo et al. (2022) found differences in emotional responses to AI-created art, with some viewers perceiving AI-generated artworks as less "sensitive" or emotionally deep than human-created pieces. Horton, White, and Iyengar (2023) similarly report that identical pieces of art are evaluated more positively when attributed to a human creator than when attributed to AI, suggesting an audience bias against AI-generated art. However, biases in judgment can be complex: other research suggests these effects may stem more from human favoritism than outright "algorithm aversion." Zhang and Gosline (2023) found that people's perceptions can actually favor human–AI collaborations under certain conditions, rating augmented human–AI works higher in quality than those created by a human or an AI alone. These boundary-conditioned effects depend on factors such as whether the art's origin (human or AI) is disclosed, the expertise level of the evaluators, the nature of the task, and the visibility of process evidence (i.e. whether viewers are shown how the piece was created).

A general pattern emerging from the literature is that HITL workflows tend to outperform HOTL pipelines on aesthetic outcomes when human creativity and judgment remain integral. When an AI system is used to assist rather than replace the human – for instance, generating suggestions that a human artist then curates, combines, or refines – the resulting artworks are often judged as more creative or of higher quality than purely machine-generated outputs. This aligns with broader findings in decision-making domains: people often exhibit algorithm aversion, a reluctance to trust automated systems' outcomes after seeing them err (Dietvorst et al., 2015). At the same time, with appropriate design and transparency, people can develop algorithm appreciation, sometimes even preferring outputs that involve algorithmic assistance (Logg et al., 2019). of art, striking the right balance appears critical. A fully automated, human-out-of-the-loop process risks "short-circuiting" important learning experiences – for example, a student might skip over learning how to compose an image if the AI does it entirely for them – and may yield work that viewers perceive as lower in authenticity or emotional resonance. Conversely, a well-designed HITL process can scaffold students' creativity by speeding up low-level tasks (like generating variations on a theme) while preserving higher-order decision making and personal input. Such an approach can also mitigate negative biases against AI involvement if it is accompanied by honest disclosure and visible evidence of the human contribution.

Despite these insights, current models and frameworks (including general tech-integration models like SAMR or even the base TPACK) remain too coarse to offer concrete guidance for art educators on these issues. Generic models classify uses of technology (e.g., substitution vs. augmentation in the SAMR model), but they do not indicate which cognitive operations are being supported or bypassed at each stage of creative work. Without a detailed mapping of AI's role at each stage, educators are often left guessing where an AI tool might help or harm the development of skills such as observation, critical reflection, or iterative refinement. Likewise, there are currently no established competency benchmarks detailing what art teachers should know and be able to do with AI, nor widely used assessment tools that capture students' creative process and ethical AI use (as opposed to only evaluating final art products). Taken together, the field lacks a coherent, actionable framework to address the gaps identified earlier. This study's theoretical framework synthesis underlines the need for an approach that explicitly links when and how to integrate AI (instructionally and cognitively) with what new teacher competencies and classroom routines are required. The remainder of this paper introduces such a framework and toolkit, and examines its initial use in practice.

METHOD

Research Design and Data Collection

This study follows a structured conceptual synthesis approach (based on scoping review principles) to build a domain-specific framework for AI-Aesthetic TPACK in art teacher education. We adopted a scoping review methodology guided by PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses – Scoping Review extension) guidelines to ensure a transparent and reproducible process (Tricco et al., 2018). No human participants were directly involved in data generation for this phase of the research; rather, our "subjects" were documents and literature sources, so institutional ethical approval was not required.

We defined the scope of our review to include published and publicly available sources at the intersection of generative AI, art/design education, and teacher knowledge/training. These sources encompassed peer-reviewed journal articles and conference papers, practitioner case reports, institutional or policy documents, and relevant methodological or assessment frameworks. The time frame was set from January 1, 2018 up to August 15, 2025 to capture the period during which diffusion-based text-to-image systems emerged and rapidly evolved. An updated sweep was conducted in late August 2025 to include the most recent publications.

We searched four electronic databases: Scopus, Web of Science Core Collection, ERIC, and Google Scholar. Search queries combined keywords in three facets: generative AI terms (e.g., "generative AI", "text-to-image", "diffusion model", as well as specific tool names like Midjourney, Stable Diffusion, DALL·E), art/design education terms (e.g., "art", "design", "studio", "visual art", "graphic", "textile", "pattern design"), and education/teacher-education terms (e.g., "education", "pedagogy", "teacher education", "teacher training", "curriculum", "assessment", "TPACK", "SAMR"). An example composite query (adapted to each database's syntax) was ("generative AI" OR "text-to-image" OR "diffusion model*" OR Midjourney OR "Stable Diffusion" OR DALLE) AND (art OR design OR studio OR "visual art*" OR graphic* OR textile* OR "pattern design") AND (education OR pedagogy OR "teacher education" OR "teacher training" OR curriculum OR assessment OR TPACK OR SAMR)

Database-specific filters were applied to focus results on English-language sources and scholarly publications (journal articles, conference proceedings, dissertations, and reviews). In addition to database results, we manually scanned the reference lists of key papers for any relevant studies that our queries might have missed. We also performed targeted web searches for institutional guidelines or policy documents on AI use in education, to incorporate practical perspectives from the field.

All retrieved records were imported into a reference management spreadsheet, where duplicates were removed. We then conducted two stages of screening: first, a title/abstract screening to exclude obviously irrelevant records, and second, full-text screening for eligibility. We included a source if it explicitly addressed art/design learning or art teacher education in relation to generative AI. We excluded sources that were purely technical (e.g. proposing a new image generation algorithm with no educational discussion) or opinion pieces lacking any empirical, conceptual, or pedagogical substance. During full-text screening, reasons for exclusion (such as "not about education" or "no relevance to art/design") were logged. The identification, screening, and inclusion process is documented in a PRISMA-ScR flow diagram (Figure 2 in Appendix), which records the number of records identified, screened out at each stage, and ultimately included.

Analysis and Synthesis

Our analysis proceeded in two iterative cycles: an inductive coding cycle followed by a deductive mapping cycle. In the inductive phase, we began by developing a qualitative codebook through exploratory reading of a subset of the included sources. We coded text segments for emergent themes related to generative AI's affordances and risks in studio learning. Examples of emergent affordance codes included "ideation acceleration", "style expansion", and "rapid iteration" – highlighting what AI made easier or faster for learners. Risk codes included themes like "homogenization of outputs", "overreliance on AI", "skill atrophy", "authorship ambiguity", and "lack of transparency". We also coded for descriptions of workflow patterns (e.g., did a scenario involve a HOTL approach with minimal human input, or a HITL approach with substantial human editing? Were prompts and iterations documented or hidden?) and any mention of alignment (or misalignment) with cognitive stages (for instance, a note that a particular use of AI might bypass the need for careful perceptual analysis, corresponding to Leder's first stage).

After refining the codebook on a trial subset, two researchers independently applied it to the full set of included sources. We achieved inter-coder reliability with Cohen's Kappa exceeding 0.70 for the top-level code categories, indicating substantial agreement on coding definitions. Discrepancies in coding were resolved through discussion and consensus. This rigorous qualitative analysis ensured that our subsequent framework construction was grounded in a broad and systematically gathered evidence base, rather than anecdotal impressions.

In the deductive mapping phase, we took the coded insights and systematically mapped them onto Leder's aesthetic cognition stages and the TPACK domains. Specifically, for each of Leder's five stages, we analyzed how generative AI was reported to function: did it scaffold learning at this stage (i.e. support or enhance the intended cognitive process), or did it shortcut/bypass the intended cognitive work? We also noted whether those effects tended to occur under certain workflow conditions (HITL vs. HOTL). For example, when considering the perceptual analysis stage, we asked: In the literature, are there instances where AI replaced students' own observation (a shortcut) versus cases where AI was used to enrich or guide observation (a scaffold)? We performed a similar mapping for each stage of the model. Alongside this, we aligned the findings with TPACK's knowledge components to articulate what knowledge or skill teachers would need to manage AI at that stage (for instance, a teacher's AI-Pedagogical Knowledge might involve strategies to prevent AI from short-circuiting students' perceptual analysis in an art lesson).

The outcome of this mapping was a detailed matrix that forms the core of our proposed framework.

This matrix specifies, for each aesthetic cognition stage: the typical HOTL uses of AI and their shortcut risks, recommended HITL scaffolds to preserve learning, examples of process evidence that should be collected at that stage, and indicative assessment criteria aligned with that evidence. This stage-by-stage mapping is presented later in the Results (as Figure 1 and Table 2), which visualizes how AI interventions can either undermine or support each stage, and what kinds of evidence and criteria can anchor a more transparent, process-focused pedagogy at each point.

Throughout the analysis and synthesis, we paid special attention to recurring practical recommendations in the literature. These often appeared in the form of design patterns or teaching tips – for example, "require an initial hand sketch before allowing AI generation" as a way to preserve students' perceptual skills, or "have students compare AI-generated variants with original concepts to practice critical evaluation." Such recurring ideas were distilled and fed into the creation of our toolkit for teacher education. In particular, they informed the descriptors and performance indicators in our competency rubric, the sections and prompts of our process-based evidence template, and the clauses of our classroom AI use and disclosure policy. In other words, the qualitative synthesis not only yielded a conceptual understanding of AI's role but also concrete instruments to implement that understanding in educational practice.

Toolkit Description

Based on the above analysis, we developed a three-part AI-Aesthetic TPACK Toolkit for art educators. The toolkit comprises: (1) a behavior-anchored competency rubric defining key dimensions of AI-A-TPACK for teachers; (2) a Process-Based Evidence Package (PBEP, 基于过程的证据包) template to document and assess students' creative process when using AI; and (3) a sample policy for classroom AI use and disclosure. Each component is described below, including its design rationale and structure. The toolkit is intended to operationalize the framework in practical settings, guiding both teachers and students toward AI use that functions as a learning scaffold rather than a shortcut.

Competency Rubric for AI-Aesthetic TPACK

One major output of our framework is a behavior-anchored competency rubric for AI-Aesthetic TPACK in art teacher education. The rubric was developed to address the identified gap in teacher competencies: art educators currently lack a clear, domain-specific model of what knowledge and skills they need to effectively integrate AI into their teaching. We distilled six core dimensions of teacher competency at the intersection of AI and art pedagogy:

AI-Technological Knowledge (AI-TK): The teacher's knowledge of generative AI tools and how to operate them. For example, understanding different image generation models and algorithms, their parameters and modes of use, and their limitations or failure cases.

AI-Technological Pedagogical Knowledge (AI-TPK): Skill in designing and orchestrating learning activities that appropriately incorporate AI. This involves knowing when and how to let students use AI in a lesson, how to scaffold its use (e.g. requiring certain human steps before or after AI use), and how to integrate AI tools into studio routines or class exercises in a pedagogically sound way.

AI-Technological Content Knowledge (AI-TCK): The ability to integrate AI into the art/design content itself. This means understanding how AI-generated content relates to art domain knowledge and techniques – for instance, using AI to explore content-specific concepts like artistic styles or color theory, and knowing how to critique or build upon AI outputs from an art content perspective.

AI-TPACK Integration: The holistic competency of aligning AI, pedagogy, and content in a cohesive manner. In practice, this means ensuring that an AI-supported task still meets the intended artistic learning objectives and fits within the curriculum's structure, and being able to "close the loop" between what students make with AI and what they are meant to learn. This dimension reflects the synthesis of TK, PK, and CK in context – essentially the essence of AI-A-TPACK.

Ethics & Policy: Knowledge of ethical guidelines, copyright issues, and the ability to model and enforce AI-related policies in the classroom. For example, teachers should know how to address questions of authorship (who "owns" AI-generated art), ensure students credit AI assistance properly, avoid biased or inappropriate AI outputs, and adhere to any school or institutional policies on AI usage. This also includes creating a classroom culture of honesty and integrity around AI use (such as requiring disclosure of AI involvement in assignments).

Assessment: The ability to assess student work in the age of AI, which includes evaluating both the final artifact and the creative process. Teachers need strategies to verify the authenticity and originality of student work, to evaluate process documentation, and to assign value to students' decision-making and effort (not just the polished outcome). This competency often involves using tools like the PBEP and rubric itself to make sure that grading considers how the student used (or didn't use) AI in academically legitimate ways.

For each of these six dimensions, we defined four performance levels: Beginner (B), Developing (D), Proficient (P), and Advanced (A). Each level is anchored by specific observable behaviors or indicators, drawn from our literature synthesis and input from expert art educators. The rubric is behavior-anchored, meaning it describes what a teacher at a given level does or can demonstrate, rather than just using abstract qualitative labels. For example, under AI-Technological Knowledge, a Beginner might "use a single generative model with default settings, treating it largely as a black box," whereas an Advanced teacher would "design a custom AI tool-chain and justify the choice of models and parameters for a given art project." Under Ethics & Policy, a Beginner might only be vaguely aware of issues (perhaps giving occasional verbal warnings about plagiarism or style theft), whereas an Advanced teacher might "coach colleagues in compliant practices" and rigorously model full disclosure of AI use in their own demonstrations.

Each cell of the rubric (dimension × level) contains a concise description of behavior. We also included a column for evidence examples – tying each dimension to the types of evidence or documentation that could demonstrate that competency in action. For instance, evidence for AI-TK might be prompt logs or screenshots showing the teacher experimented with different models and settings; evidence for AI-TPK might include lesson plans or classroom observation notes showing how the teacher integrated AI into an activity; evidence for Ethics & Policy might include completed AI disclosure forms or examples of the teacher explicitly addressing AI ethics with students. The rubric is designed to be scored holistically: we assigned indicative weightings to each dimension (summing to 100%) to reflect their importance. In our current iteration, the Integration and Assessment dimensions carry slightly higher weight (e.g. ~20% each) because of their critical role in linking AI use to learning outcomes, while dimensions like AI-TK, AI-TPK, AI-TCK, and Ethics/Policy might carry around 15% each. These weights, along with cut-off scores for overall competency levels (for example, a total score ≥ 75% indicating Proficient overall), can be adjusted or calibrated in future studies. For now, they serve as a guideline.

To illustrate, Table 1 (see Appendix B-2 for the full rubric) shows a simplified excerpt of two dimensions with behavioral anchors at different levels:

Table 1. Example rubric segments for two dimensions (AI-TK and Ethics & Policy) across selected proficiency levels.

Dimension	Beginner (B)	Proficient (P)	Advanced (A)
AI-Technological Knowledge (AI-TK)	Uses one AI tool with default settings; limited understanding of how outputs are generated.	Experiments with multiple AI tools and settings; understands basic model differences (e.g. resolution, style strengths).	Customizes AI workflows; combines models or writes code to fine-tune outputs; explains model choices and limits for specific art tasks.
Ethics & Policy	Vaguely aware of plagiarism concerns; might warn students not to "cheat" with AI but gives no formal guidelines.	Implements class rules for AI use (e.g. requires citation of AI assistance); addresses obvious ethical issues when prompted.	Proactively teaches about AI ethics and copyright; models full AI-use disclosure in demos; enforces a formal AI policy and mentors others in ethical best practices.

The rubric is intended to be used in multiple ways: (a) as a self-assessment or formative assessment tool for pre-service or in-service teachers to gauge their own growth in learning to integrate AI; (b) as a curriculum design guide for teacher educators, to ensure that training programs address all the dimensions (for example, including explicit training on ethical AI use, not just technical skills); and (c) as a research instrument, providing an operational definition of AI-related teaching competency that could be used in future studies (for instance, to measure the impact of a professional development intervention on teachers' AI integration skills). In developing the rubric, we consulted existing teacher competency standards and behavior-anchored rating scales to ensure our descriptors were clear and actionable. By translating abstract ideas of "AI in education" into concrete behaviors tailored for art educators, the rubric fills a critical void and offers a tool for both guiding and evaluating teacher preparation in this new domain.

Process-Based Evidence Package (PBEP) Template

To address the lack of process-focused assessment in AI-assisted artmaking, we developed a Process-Based Evidence Package (PBEP) template as part of the toolkit. The PBEP is essentially a structured portfolio or log that students (or teachers, in training scenarios) fill out to document the process of creating an artwork when using generative AI. The goal of the PBEP is to make the normally hidden creative process visible and assessable, rather than judging only the final artifact. Drawing on recommendations from both our literature review and emerging best practices in digital art education, the PBEP template comprises several components (Appendix A provides the full blank template and an example):

A1. Prompt Iteration Log (required): A table for recording each major prompt iteration and AI generation step. For each entry, the student notes the date/time, the AI model (and version) used, key parameter settings, the exact prompt given (or a summary of it), and the goal of that iteration along with a brief outcome note. This log encourages reflection on how the prompt and settings were tuned over time and creates traceability in the creative process. Example entry: Date – 2025-03-12; Model – Stable Diffusion XL 1.0; Parameters – CFG=7, steps=30; Prompt – "bronze sculpture of a dancer, modern minimalist style"; Goal – refine figure's pose; Outcome – iteration #3 had a better pose but background was cluttered, will adjust prompt for background. A snippet of this log is illustrated in Table 2 below.

Table 2. Sample entry from a student's Prompt Iteration Log (PBEP Component A1).

Date	AI Model (Version)	Key Parameters	Prompt	Goal	Outcome
2025-03-12	Stable Diffusion XL 1.0	CFG=7; steps=30	"bronze sculpture of a dancer, modern minimalist style"	Refine figure's pose	Iteration #3 had a better pose, but background was cluttered; will adjust prompt to fix background.

A2. Human Modification Sheet (required): A section to document any manual edits or hybrid steps taken after obtaining AI outputs. Here, the student records each significant human-driven modification to the work—such as compositing multiple AI-generated images in Photoshop, adjusting layouts or color balances, painting over certain areas, adding textures, etc. For each step, the student notes what was done, what tool or medium was used, and the rationale (e.g., "Combined two AI outputs for background and foreground; painted highlights on the main figure using Procreate to enhance lighting; rationale: to integrate multiple ideas and add hand-crafted emphasis on lighting."). This sheet highlights the human contributions and craftsmanship involved in the final piece.

A3. Observation & Dimension Sheet (required): This component fosters engagement with the early cognitive stages (perceptual analysis and explicit classification) without relying on AI. Students include brief observational notes and ratings of foundational visual elements or design principles relevant to their project (such as line quality, color harmony, balance, rhythm, etc.), typically on a simple scale (e.g. 1–5 for each element). They also have space to insert a small grid of sketches or thumbnails (for instance, a 3×3 grid for initial idea sketches drawn by hand) and to cluster visual references or style images. The idea is to ensure the student engages in non-AI-supported observation and concept development before or alongside using AI, anchoring their creative process in traditional skills and personal artistic intent.

A4. Asset & License Register (required): A list or table to inventory any third-party assets or data used in the project, such as textures, base images, or datasets, along with their sources and license information. With generative AI often drawing from large training datasets (and students frequently using online reference images), it's crucial to track anything not original to the student. This register teaches students to respect copyright and open-license practices, and it provides transparency about which components of the work are original versus borrowed or AI-generated.

A5. AI Use Disclosure Checklist (required): A checklist for explicitly documenting the details of AI usage in the project. It prompts the student to specify which AI model(s) and version(s) were used, on what dates, with which settings; which parts of the final work were AI-generated versus significantly human-modified; what post-processing tools or steps were applied to AI outputs; and includes a declaration for the student to sign, confirming they have disclosed all AI assistance and complied with the course's AI use policy. In essence, this mirrors what an honest "methods section" of their artistic process would look like. By institutionalizing disclosure, it normalizes transparency and removes the temptation for students to hide AI use. It also provides teachers with context to better assess the work.

A6. Peer Review Form (optional): If peer critique is part of the class, this form allows a peer (or multiple peers) to review the transparency and quality of the process, not just the final product. Peers can rate aspects like how traceable the process is (i.e., can they follow what steps the original student took?), the depth of human contribution (did the work show significant human creativity and editing, or was it mostly AI output?), the aesthetic coherence of the final piece (does it all come together as a strong artwork regardless of how it was made?), and compliance/honesty (does it appear the student followed all rules and disclosed properly?). The peer reviewer also provides comments. Including peers in evaluating process can instill a culture of accountability and reflection. We mark this component optional because not every context will use peer review, but we provide the template for those that do.

Each component of the PBEP corresponds to evidence that can be used in assessment and aligns with stages of the creative process. We also designed a simple scoring guide for the PBEP (suggesting, for example, that completeness of documentation might count for 20%, traceability and clarity of process 20%, depth of human contribution 30%, disclosure & compliance 20%, and peer review feedback 10%, totaling 100 points). These weights emphasize that simply using AI is not enough – the student must show thoughtful iteration and honesty about their process. In a classroom implementation, the PBEP would be submitted alongside the final artwork, and teachers could grade it using the provided criteria or an associated rubric. (In fact, our teacher competency rubric references these evidence pieces on the teacher side, e.g. expecting teachers to check students' disclosure forms as part of their assessment competency.)

The full blank PBEP template is included in Appendix A, which educators can directly adopt or adapt for their classes. By capturing and formalizing process evidence, the PBEP shifts some focus of assessment from solely the artifact to the learning process, encouraging students to engage in reflective and responsible use of AI.

AI Use and Disclosure Policy

The third component of our toolkit is a Sample Policy for Classroom AI Use and Disclosure (see Appendix B). While the rubric and PBEP template are aimed at guiding individual behavior and assessment, the policy provides a course- or program-level framework to ensure all participants are on the same page regarding acceptable AI use. The need for such a policy became evident from our gap analysis: many educators lack clear policies bridging ethical principles with day-to-day practice, beyond generic academic integrity statements.

The policy document is structured into several sections:

Scope and Purpose: It states that the policy applies to all generative AI use in the course, and frames the purpose as harnessing AI as a learning scaffold while maintaining academic integrity and respecting intellectual property. This section sets a positive tone – AI is not banned outright, but its use is guided and transparent – and aligns the policy with broader institutional values like honesty and learning.

Encouraged Uses of AI: The policy explicitly lists examples of AI use that are pedagogically beneficial and allowed. These might include using AI for brainstorming and ideation (e.g. to generate rough ideas or concept sketches in the early stages of a project), generating contrast sets or variations to explore the boundaries of a design concept, using AI for local refinement (for example, applying an AI tool to try multiple texture options for a background, which the student then integrates and adjusts), and using AI's process-recording features to document steps (for inclusion in the PBEP). By enumerating encouraged uses, the policy actively educates students and staff on what productive AI use looks like, reinforcing the idea of AI as a scaffold (not a shortcut).

Mandatory Disclosure: A critical section that mandates transparency. It requires that for any assignment where AI is used, the student must clearly disclose what was done with AI. This includes specifying the model and version, the date of use, all prompts or inputs given, any key parameter settings changed from default, any post-processing tools applied to AI outputs, and which parts of the submission were AI-generated or AI-modified. Essentially, this mirrors the information captured in the PBEP's disclosure checklist (Appendix A5). The policy statement might read, for example: "Students must include a completed AI Disclosure Checklist with every submission involving AI. The checklist must detail the model name and version (e.g., Midjourney v5, DALL·E 3), the date of use, all prompts or instructions given to the AI, any non-default parameters used, any edits or additional tools used on AI outputs, and which portions of the final work were generated by AI." By institutionalizing disclosure, the policy normalizes honesty and ensures instructors have the context needed to fairly assess the work.

Prohibited Uses of AI: This section clearly delineates what is not allowed. Examples include submitting AI-generated work without any meaningful human modification (i.e. you cannot simply type a prompt and submit the raw AI output as your "artwork"), using AI to replicate someone else's style or a specific existing artwork in a way that infringes on copyright or violates academic integrity, using any AI tool or service that is explicitly disallowed by the course or institution, and providing false or incomplete disclosure of AI use. By listing prohibited actions, the policy guards against common failure modes – like over-reliance on AI to do the work for the student, misrepresentation of authorship, or violating others' intellectual property. Essentially, it draws a clear line between using AI as a creative aid versus using AI in a way that undermines learning or ethical standards.

Assessment and Accountability: The policy ties into the course's grading system and consequences for violations. It specifies that failures to follow the policy (e.g., not disclosing AI use, or using AI in a forbidden manner) will impact grades and could trigger academic misconduct procedures. For instance, not submitting the required disclosure or PBEP might result in an automatic deduction (say, up to 30% of the assignment's points, since process accountability is integral to the assignment). Submitting work that is essentially unedited AI output can receive a major penalty or a zero for that portion, with a requirement to redo the work. Use of unauthorized material or serious misrepresentation (like lying on a disclosure form) would be handled under the institution's academic integrity policy, potentially resulting in failing the assignment or course, or other disciplinary measures. By explicitly linking the policy to grading and enforcement, we ensure the rules have "teeth" and are not merely aspirational.

Overall, the sample policy (provided in full in Appendix B) can be adapted by instructors or institutions to fit their needs. Its presence at the start of a course is intended to set the tone that AI is a welcomed tool with conditions — students learn that how they use AI, and how transparently they use it, will matter just as much as the final artwork they produce. This helps cultivate an environment where AI is seen as part of the learning process rather than a shortcut to circumvent it.

By combining the competency rubric, PBEP template, and policy, our toolkit addresses the gaps in teacher guidance, student process accountability, and class policy around AI use. The next sections present the findings from our stage-based analysis that underpin this toolkit and a preliminary evaluation of the toolkit's use in practice.

RESULTS AND DISCUSSION

Stage-by-Stage Mapping of AI Interventions (Scaffolds vs. Shortcuts)

Our stage-by-stage analysis of the literature yielded a clear pattern of how AI can either scaffold or shortcut student learning at each stage of aesthetic cognition, depending on the workflow design. Table 3 summarizes these findings in a tabular format, mapping each stage of Leder's model to typical AI uses under a HOTL approach (and the associated risks) versus recommended HITL scaffolds, along with examples of process evidence and indicative assessment criteria for each stage:

Table 3. Mapping AI interventions across aesthetic cognition stages under HOTL and HITL workflows.

Aesthetic Stage	HOTL usage (risk)	Harm/Risk	HITL scaffold (recommended)	Process evidence	Indicative criteria
Perceptual analysis	One-shot AI generation replaces seeing (skips direct observation)	Skips foundational noticing of details	"First human, then model" – e.g. require students to do quick hand sketches and note key details/colors before using AI	Sketch sheets; initial prompt draft	Depth of observation; alignment of AI prompt with observed details
Implicit integration	AI imitates a known style	Homogenization of style;	Human-driven style exploration then AI contrast –	Style mood boards or	Originality of contrasts; degree of

	(style cloning)	superficial analogies	e.g. students manually cluster example artworks into style families, then use AI to generate boundary-pushing variants between clusters	clusters; AI-generated contrast images	expansion beyond familiar styles
Explicit classification	AI auto-labels or categorizes images (offloads concept formation)	Offloads cognitive work of classification to AI	Student articulates categories first, AI only used to generate exemplars – e.g. fill out a “dimension sheet” of style rules, then have AI produce examples that fit those rules for discussion	Student-created dimension sheet; AI exemplar set	Quality and clarity of student-defined classification rules; accuracy of AI exemplars in illustrating the rules
Cognitive mastering	End-to-end AI generates final composition (no human iteration)	Collapses iterative problem-solving; student does minimal refinement	Two-step human–AI remix: e.g. use AI for an initial draft, then require human editing and only then allow selective AI re-generation for specific parts	Layered image files; edit logs (showing human edits)	Iteration depth; evidence of problem-solving in edits; improvements made between AI draft and final piece
Evaluation	Judged only on final AI-produced artifact (process invisible)	Overestimation or misattribution of quality; learning process unexamined	Dual-track review: blind critique of the artifact plus review of process evidence (PBEP) using rubric	Completed rubric scores; disclosure checklist	Balanced evaluation (product quality weighed alongside process integrity and effort)

Each stage lists common HOTL usage patterns (and their pitfalls) contrasted with recommended HITL scaffolding strategies. The rightmost columns illustrate the types of process evidence to collect (see PBEP in Appendix A) and suggest criteria aligned with our rubric for evaluating student performance. The overarching principle is that HITL approaches explicitly link task design, evidence visibility, and assessment criteria to ensure AI acts as a scaffold rather than a shortcut.

As Figure 1 shows, at the Perceptual analysis stage, a HOTL approach might let students immediately generate an image with AI from a prompt, effectively bypassing direct observation and sketching. The risk is that students do not engage in seeing and processing visual details themselves – they outsource the initial visualization to the AI. The HITL recommendation is to enforce a “first human, then model” rule: for example, in a classroom a teacher might require students to spend the first 10–15 minutes closely observing a reference object or scene and making a few quick sketches (noting key shapes, colors, textures), only after which can they input a text prompt to generate an image. Process evidence like sketch sheets or a written first-draft prompt (informed by their observation) can verify that this human-led step happened. The assessment criteria at this stage would then include how well the student's AI prompt or generation reflects careful observation – in other words, did the student notice enough detail on their own to guide the AI effectively?

At the Implicit Memory Integration stage, a naïve use of AI is to prompt an AI system to generate in a specific artist's style (e.g., "generate an image in the style of Van Gogh"). This often results in style cloning—work that imitates known styles without the student internalizing or personalizing the underlying traits. The risk is homogenization and superficial analogies: students may converge on popular or clichéd styles and fail to connect them to their own expressive memories or insights. The HITL scaffold is to require a human-led style exploration first: have students manually gather and cluster exemplars (e.g., a mood board or a style taxonomy), discuss what they personally notice and recall about those styles, and only then use AI to produce contrast images that mix or push their boundaries. Process evidence may include photographs of the clustering activity, notes on personal connections, and the resulting AI-generated contrasts. An indicative criterion is the originality and insightfulness of the contrasts—did the student use AI to move beyond obvious tropes, showing deeper integration of a personal perspective?

For the Explicit classification stage, a problematic approach would be to use an AI tool to automatically label or categorize images (for example, using an AI image recognition model to tell the student the style or elements present, instead of the student identifying those). This offloads the intellectual work of concept formation and categorization to the AI, bypassing a valuable learning opportunity. The scaffolded approach is to have students themselves define classification schemes or rules first – for instance, filling out a dimension sheet where they list key features or rules that distinguish different styles or techniques they're studying. Only after they have made their thinking explicit, they might use AI to generate example images that fit those categories or rules, which can then be critiqued. Evidence for this would be the student's written notes about their classification criteria and the set of AI-generated exemplars they curated. The criteria for evaluation could include the quality of the student's classification logic (did they identify meaningful dimensions?) and the accuracy or appropriateness of the AI exemplars in illustrating those concepts.

At the Cognitive mastering stage – where iteration, refinement, and problem-solving typically happen – a HOTL shortcut would be if a student leans on the AI to handle the entire iteration process. For example, the student might keep hitting "generate" or slightly tweaking the prompt until the AI output looks acceptable, without engaging in any reflection or substantial revision of their own. This means the AI is doing the heavy lifting of resolving problems or exploring options, and the student isn't practicing those skills. The recommended HITL design is to enforce a two-step human–AI remix: the student uses AI to get an initial draft or set of ideas, then must step away from AI to do a round of human edits or recombination (e.g. pick the best elements from different AI outputs and manually merge or redraw parts, or critique the draft and make changes by hand), and only after making a substantial human contribution can they optionally use AI again for targeted refinements (like regenerating just one segment of the image where they want fresh ideas). This approach forces the student to engage in the cycle of idea → draft → critique → refine, with AI as a partner rather than the sole creator. Process evidence could include layered image files or an edit log showing the sequence of human modifications. Criteria might look at the depth of iteration (how many meaningful changes were made, how the work evolved from initial to final) and evidence of problem-solving in those edits (for example, did the student clearly address issues from the first draft in the final outcome?).

Finally, at the Evaluation stage, a fully HOTL scenario might treat an AI-assisted artwork just like any other and grade it solely on the final product, or worse, if the AI's involvement is unknown, the teacher might unknowingly over-credit the student or be biased if AI use is suspected. This is the "invisible process" risk: important context is missing, so either the student might get undue credit for what the AI did, or conversely a stigma might be applied to AI usage. The HITL approach for evaluation is to implement a dual-track evaluation: one evaluation focuses on the artifact's aesthetic quality (ideally done "blind" to how it was made, to judge it on its own merits), and another evaluation focuses on the process documentation (using tools like the PBEP and rubric to assess creativity, effort, and integrity in how the piece was produced). By combining these, the teacher (or peer reviewers) can form a more holistic assessment. For instance, a piece that is modest in its final appearance but shows a rich, thoughtful process would receive credit for learning and effort, whereas a visually stunning piece that was mostly AI-generated with very little student input would be marked down on process, even if the product is attractive. Our rubric explicitly weights both product and process, meaning an educator could, for example, assign 50% of the grade to the final artifact quality and 50% to the process quality (or some similar scheme). The process-related criteria might include things like transparency (did the student fully disclose and document?), originality of contribution (what did the student themselves add?), and adherence to policy.

To illustrate how these stage-aligned strategies can play out in practice, we implemented a 90-minute class workshop as a model lesson incorporating all stages with corresponding scaffolds (this also served as a trial of the toolkit, described later). The session was structured as follows: (1) 15 minutes – Guided observation and hand sketching: Students closely observed a physical object and made quick sketches, noting key visual details (perceptual analysis stage, without AI). (2) 20 minutes – Style exploration with clustering and dimension sheets: Students reviewed small printouts of artworks in various styles, clustered them into groups, and identified distinguishing features, while reflecting on personal connections (implicit integration and explicit classification stages, human-driven). (3) 20 minutes – AI-assisted generation of contrast sets: Students then used a text-to-image AI tool to generate a few images that intentionally mixed or contrasted the styles they had been exploring. They selected a couple of interesting AI results and explained why those stood out (transitioning from classification to cognitive mastering). (4) 20 minutes – Human remix and refinement: Using the AI outputs as a starting point, each student edited one image—some combined elements from multiple AI outputs, others painted over parts or adjusted composition—to develop a more polished piece, applying their own creative decisions (cognitive mastering via human-led iteration). (5) 15 minutes – Dual-track critique: Half the class conducted a peer blind critique, exchanging final images without process info and giving feedback on the artwork's impact (evaluation of product), while the other half exchanged PBEP logs and reviewed each other's processes. Then they switched, and finally discussed both the products and the processes, using a simplified version of our rubric to guide feedback (evaluation stage with full transparency).

This workshop sequence demonstrated that even in a single class session, it is feasible to engage each stage of aesthetic cognition meaningfully with AI. The toolkit components were integral: the rubric provided criteria for peer feedback during critique; the PBEP logs were the basis of process-oriented discussion; and the policy was referenced to remind students to be honest about AI usage. Students reported that having to document and explain their process made them more aware of their own decision-making ("I found myself planning my next prompt more carefully instead of just trial-and-error"), and teachers noted that the critiques were richer because of the explicit process evidence available. This example suggests that a stage-aligned approach to AI integration can be practical even within typical class time constraints, not only in long-term projects.

Rubric Usability and Consistency Check

To gauge the usability and consistency of the AI-A-TPACK competency rubric (Table 1 in the toolkit) in practice, we conducted a small-scale trial with a group of art teacher educators. We recruited three experienced art instructors (each with 8–15 years of teaching experience) who were not involved in the toolkit's development to serve as test raters. They participated in a 90-minute pilot workshop on 18 June 2025, where we introduced and calibrated the rubric and then had them apply it to a sample of student work.

The workshop followed a calibrate–score–review sequence. First, the three instructors spent ~15 minutes in a guided calibration session: we introduced the six rubric dimensions and walked through one example case as a group, discussing what different performance levels might look like for that case. This helped establish a shared understanding of the criteria at each level (from Beginner through Advanced). Next, each instructor independently applied the rubric to assess a sample Process-Based Evidence Package (PBEP) from a hypothetical student project. They were given about 20 minutes to review the student's final artwork and the complete PBEP documentation, and to assign rubric levels for each of the six dimensions without consulting each other. Immediately after scoring, we facilitated a ~15-minute review discussion in which the instructors compared their ratings, explained their reasoning, and worked through any discrepancies in their judgments.

Inter-Rater Reliability and Discrepancies: The preliminary scoring results from this exercise showed encouraging consistency. Out of the 6 rubric dimensions, all three instructors agreed exactly on the performance level for 4 dimensions in the sample case, indicating strong initial alignment on those aspects of teacher competency. For the remaining 2 dimensions, scores differed by at most one level between raters. For instance, two raters marked the sample student as "Proficient" in AI-Technological Content Knowledge while one thought it was closer to "Developing," and similarly they differed slightly on the Assessment dimension. We calculated Cohen's Kappa as an index of inter-rater reliability: at the individual dimension level, κ ranged from approximately 0.58 to 0.73, and the overall agreement across all dimensions was κ ≈ 0.65 (which is generally interpreted as moderate to substantial agreement given the small number of raters). Notably, the AI-TCK and Assessment dimensions had the lowest raw agreement, whereas the other dimensions (AI-TK, AI-TPK, AI-TPACK Integration, Ethics & Policy) had complete agreement among all raters on the sample.

Through the facilitated discussion, we discovered the primary cause of those minor divergences. In both cases, it was about how strictly to interpret certain criteria. For example, on the Assessment dimension, one instructor penalized the sample student for not crediting a third-party texture library they had used (considering this a significant omission under "process transparency"), whereas another instructor hadn't noticed that omission or didn't weigh it as heavily. When the group discussed this, they reached a consensus that failing to disclose a significant external asset (even a free-to-use resource) should be considered a policy violation and warrant a deduction in the process score. They agreed to treat that scenario as a serious issue in scoring. We documented this decision as a precedent for future rubric use (essentially adding a note to the rubric guidelines: "all third-party assets must be credited or it impacts the Ethics/Policy or Assessment dimension score"). Apart from clarifications like this, the instructors reported that the rubric descriptors were sufficiently clear and concrete. One rater commented, "I could almost quote the student's evidence that matched the descriptors – it was clear what counted as, say, Proficient vs Developing in each category." Another noted that having the evidence examples column to refer to made the scoring more objective: "It guided me to look for the prompt logs or the disclosure form when assessing Ethics & Policy, for example, which made my judgment more grounded in facts." All agreed that the brief calibration exercise prior to scoring was critical in achieving the high agreement we observed.

This transparent reporting of even the small points of rater disagreement (a "negative" finding of sorts) actually strengthens the credibility of our results – it shows that we identified and addressed potential ambiguities in the assessment tool. We have effectively started a "casebook" of such issues and resolutions (two examples are provided in Appendix C), which can guide future users of the rubric.

Overall, this initial trial indicates that the rubric is both usable and yields reasonably consistent evaluations among different raters after minimal training. For high-stakes use (like formally evaluating a student teacher at the end of a program), we would recommend a more extensive norming session with the rubric and possibly using multiple raters to ensure fairness. But as a formative tool (for self-assessment or guiding mentorship feedback), even a single instructor using it would likely find it provides a structured, comprehensive lens on a teacher's competencies, given its coverage of key areas.

In sum, this preliminary evidence of consistency addresses our second research objective: it provides confidence that the toolkit's rubric can be applied in real classroom contexts without undue ambiguity, even within the short timeframe of a workshop or a single class review session.

DISCUSSION

A key insight from our synthesis and trials is that whether AI functions as a scaffold or a shortcut in learning is not inherent to the tool itself, but contingent on how it is used and structured within the learning activity. Several boundary conditions emerged:

1. Human Effort and Edit Depth: AI acts as a scaffold when the human student's effort is substantive and traceable. Whenever the student invests meaningful effort – whether in carefully crafting the prompt, adjusting parameters thoughtfully, or significantly editing the AI outputs – the AI tends to serve as a partner in learning rather than replacing the learning. In contrast, if a student simply accepts the first AI output uncritically or with only trivial tweaks, the AI is doing too much of the cognitive work, effectively becoming a shortcut. Our rubric and PBEP explicitly emphasize edit depth and rationale to encourage substantive engagement. For instance, the rubric's iteration criteria and the PBEP's modification log both push students to go beyond one-shot generation and into meaningful transformation of AI outputs.
2. Process Documentation: We found that when prompts, parameters, and decision rationales are recorded and made visible (as required by the PBEP and policy), students approach AI use more metacognitively – that is, they think about their thinking. The act of documentation turns each AI interaction into a point of reflection ("Why am I doing this? What did it yield? What should I try

next?"). This tends to make AI a scaffold because the student is actively learning from each step and considering alternatives. Conversely, when AI use is undocumented or hidden, it invites shortcut behavior. Students might engage in rapid trial-and-error prompting without reflection, or even use AI in prohibited ways, hoping no one will notice. The requirement to document process serves as a form of accountability that nudges usage toward learning-oriented patterns.

3. Peer and Mentor Feedback on Process: When critique or feedback sessions incorporate process evidence and not just final products, it closes the loop and holds students accountable for their approach, not only the outcome. In our workshop, for example, peers had to comment on whether the process was transparent and whether the student made meaningful contributions. This clearly signaled to students that just producing a pretty picture isn't enough – how they got there matters to their peers and instructors. If no one ever looks at process, students quickly learn that only the output "counts," and some will logically lean on shortcuts if those yield decent outputs. By contrast, if teachers and peers are routinely asking, "Show me how you did this," students are incentivized to engage with AI in a way they can explain and stand behind.
4. Course Policy and Culture: A course policy that mandates disclosure and emphasizes learning over product sets a scaffold-oriented culture from day one. The presence of our sample policy was noted by the pilot teachers as an important signal to students: it explicitly tells them that using AI is acceptable only of learning and honesty. In an environment with no policy or a very lax stance, students might either feel they should hide AI use (leading to secret shortcuts and anxiety) or conversely think they can use AI without limits or personal effort (leading to open shortcuts). A clear policy with enforcement provisions nudges behavior toward the intended scaffolded use by establishing norms and consequences upfront. It essentially externalizes the teacher's expectations so students are not guessing where the line is.

These factors echo findings in broader human–technology interaction research. For instance, algorithm aversion (Dietvorst et al., 2015) can be mitigated by giving people more control or understanding of the algorithm. In our context, when students have control (via HITL workflows) and understanding (via documentation and reflection) of the AI, they integrate it more constructively. Similarly, what some call "human favoritism" or a bias to value human contribution (Zhang & Gosline, 2023) is intentionally built into our approach: we require a human contribution at each stage, so the final outcome is genuinely a human–AI collaboration, not just AI output with a human's name on it. Interestingly, we observed that the same AI tool or action can be either a scaffold or a shortcut depending on the learner's level and intent. For a novice lacking foundational skills, having an AI automatically handle a task (say, perspective drawing) might shortcut their learning of that skill. But for an advanced student who already understands perspective, using AI to quickly apply it could be an efficient scaffold that lets them focus on higher-level composition decisions. This implies the framework should be applied flexibly and educators must exercise judgment: for example, a teacher might allow more AI assistance for advanced students exploring complex projects, but require more manual work from novices who need to build fundamentals. In teacher education, we emphasize this adaptive use – part of AI-TPK is knowing your students and deciding where their line between scaffold and shortcut lies.

In summary, AI does not intrinsically know whether it's scaffolding or short-circuiting learning – educators design the context that determines that. Our framework's value lies in making those conditions explicit and providing tools to manage them (e.g., rubric criteria for sufficient human editing, required process logs, disclosure norms, etc.). By tuning these levers, teachers can tilt AI usage toward being a catalyst for learning rather than a crutch.

Integrating the Toolkit into Teaching Practice

Bringing together the framework and toolkit, we envision a restructured approach to art teacher education and classroom practice in the era of AI. The components we developed are meant to function in unison: the stage mapping informs curriculum design, the rubric guides teacher development and assessment, the PBEP template structures student activities and evaluation, and the policy creates an environment of integrity and clarity.

In a teacher preparation program, for example, coursework can be explicitly aligned with these components. A unit on "AI in Visual Analysis" (covering the perceptual stage) might train pre-service teachers to use the "first human, then AI" strategy in their future classrooms. An assignment could have them design a lesson plan where students do observational drawing before any AI imaging. The rubric's AI-TPK and AI-TCK dimensions would come into play in evaluating their lesson designs, and the instructor of the course could use rubric language in feedback (e.g., "Your lesson plan is at a Developing level for AI-TPK because it lets AI take over too early; how could you modify it to scaffold human observation first?"). Another module on "Ethical AI Practice" could revolve around the policy: teacher candidates might be tasked with adapting the provided AI use policy for a hypothetical school scenario, and practicing filling out the disclosure forms for sample projects. This addresses the Ethics & Policy dimension of the rubric in a hands-on way.

In actual K-12 or college art classes, implementing the toolkit means that when teachers integrate an AI-based assignment, they do a few things differently: they introduce the policy to students (so expectations and consequences are clear from the start), they require the PBEP or some process documentation as part of the assignment deliverables (ensuring students document their process, not just turn in a final image), and they employ a rubric – or a simplified version of it appropriate for students – to evaluate the work, explicitly giving weight to process and ethical compliance. The "dual-track" evaluation approach mentioned earlier can be operationalized by splitting the grading criteria into those focusing on the final product and those focusing on the process. Some teachers might combine this into a single score (with sub-scores that students see for product vs. process), while others might even give separate grades for the artwork and the process portfolio, depending on their philosophy and school grading policies.

One practical advantage of having a well-defined rubric and evidence structure is the potential for consistency and fairness in grading. Art assignments can be notoriously subjective to grade; however, if students provide concrete process evidence and teachers have predefined criteria (e.g., "did the student show at least X iterations," "did they credit all sources," "does the final piece reflect personal decision-making?"), it helps different instructors remain aligned in their evaluations. This also helps in communicating expectations: students know that to score well, they cannot just turn in a beautiful image – they also need to show how they arrived there and that they followed all guidelines. This transparency can reduce conflicts or confusion over grading, as both parties can point to documentation and specific rubric descriptors.

Our pilot workshop, though small in scale, provided a microcosm of how this toolkit can change classroom dynamics. We observed strong student engagement and even some shifts in mindset. Initially, a few students admitted they thought using AI might let them "skip ahead" in an art task or avoid learning something difficult (for example, not bothering to learn how to draw hands because "the AI can do it"). After the workshop, they reported a better understanding that AI is a tool that still requires skill and intent. One student reflected, "Because I had to log what I was doing and why, I actually found myself planning my next prompt more carefully instead of just trial-and-error." This indicates a deeper learning outcome: the toolkit not only measures but actively promotes reflective practice.

From the teacher's perspective, the toolkit can also transform their role. Teachers become designers of scaffolded workflows – they actively think about which stage of the creative process they want to emphasize at a given moment and how AI could either support that or hinder it. They also become coaches on process and integrity, not just evaluators of finished art products. Adopting the toolkit encourages teachers to openly discuss questions like: How do we decide if using AI at this step is helping you learn or doing the work for you? What does it mean to be original in an age of AI assistance? These discussions tie directly into art and design pedagogy topics (authorship, creativity, developing an artistic voice) but in a very contemporary way.

Challenges and Limitations

While we advocate for a structured approach, we caution that like any framework, there is a risk of overemphasis on "compliance" if the toolkit is misused. For example, students (or teachers-in-training) might become overly focused on checking all the boxes – filling out logs, disclosing everything – just to get a good score, without truly internalizing the deeper purpose (which is to enhance learning and critical thinking). This type of performative compliance has been noted in other educational contexts where rubrics or checklists are introduced. To mitigate this, teacher educators and instructors should continuously highlight the why behind each component. For instance, remind students that the reason we document process is to become more mindful artists and to be able to learn from mistakes, not merely because it's a requirement. Embedding reflective prompts ("What did you learn in this iteration?") in the PBEP, or having periodic class discussions about experiences using the toolkit (where students share not just what they did, but what they thought about it) can help ensure the toolkit is a means to an educational end, not an end in itself.

Another consideration is the adaptability of the framework across different cultural or educational contexts. Our use of Leder's aesthetic stages and certain design patterns might be grounded in a Western-centric model of art education, which often emphasizes individual creative process, originality, and explicit reflection. In some other educational traditions, there might be a stronger emphasis on apprenticeship, mimicry of masters, or community-oriented creation. Those contexts might require tweaking the framework – perhaps combining or re-weighting some stages, or adjusting the rubric descriptors to local values. For example, if a curriculum values technique replication as a step (copying a master's work to learn technique, which is common in some classical art training), AI might be used differently there (maybe as a source of "master style" examples) and the evaluation of process might differ. Similarly, in art forms outside the typical Euro-American canon, the stages of creative process might not map neatly to Leder's categories. Future research and collaboration with educators in diverse settings will be important to test how the AI-A-TPACK toolkit works in those environments and what modifications are needed.

Finally, we must consider the long-term impact on learners. Our framework assumes that by being forced to articulate their process and by seeing AI as a partner rather than an answer machine, students will develop better creative thinking and meta-cognitive skills. This is an implicit hypothesis that should be verified. Over time, do students who learn with this framework show greater growth in, say, their ability to generate original ideas or to critically evaluate art (with or without AI)? Do they become more adept at using AI responsibly and creatively when they are on their own? We believe so, but longitudinal research or case studies following students or new teachers over a year or more would greatly strengthen the argument. Additionally, as AI tools evolve (becoming perhaps even more powerful or more user-friendly), the strategies might need updating. The ethical norms might shift as well (for example, if legal standards for AI-generated content change, policies will need updates).

In conclusion of this discussion, integrating AI into art education demands a rethinking of pedagogy, assessment, and teacher preparation. Our proposed framework offers a structured path forward, aiming to ensure that generative AI becomes a catalyst for deeper learning and creativity rather than a shortcut that bypasses skill development. The initial trials and examples we provided are encouraging: they show that with clear frameworks and tools, both teachers and students can navigate this new landscape in a productive and ethically responsible way. However, flexibility, reflection, and continuous refinement will be key. The conversation between educators, students, and researchers must remain open as we collectively learn how to best blend human creativity with artificial intelligence in the art classroom.

CONCLUSION

This study proposes AI-Aesthetic TPACK as a stage-aligned, practice-ready framework for redesigning art teacher education in the age of generative AI. By mapping Human-in-the-Loop and Human-out-of-the-Loop workflows onto the specific stages of aesthetic cognition, we clarify where AI can scaffold artistic learning and where it risks short-circuiting essential cognitive processes. We have operationalized this framework into a concrete toolkit comprising a competency rubric for art educators, a process-focused evidence template (PBEP) for student work, and a model policy for AI use and disclosure in the classroom. Together, these contributions offer a coherent pathway for teachers to integrate AI in ways that maintain transparency, encourage creative process, and uphold academic integrity.

The work presented here is primarily a conceptual and design-based contribution, laying a foundation for future empirical validation. Our initial pilot of the rubric with multiple raters demonstrated that the framework can be applied consistently with minimal training, and our classroom workshop example illustrated its practical feasibility and positive influence on learning behaviors (like increased reflection and honesty about AI use). These early indications are promising. We recognize, however, several limitations and avenues for further inquiry:

First, while the rubric and toolkit components were informed by a broad literature base and iterative expert feedback, they have not yet been widely tested across different educational programs or demographic contexts. Future research should formally evaluate the rubric's reliability and impact in various teacher education settings. For instance, do teachers who train with the AI-A-TPACK approach later demonstrate more effective AI integration in their own classrooms compared to those who don't? Quasi-experimental studies or larger-scale implementations could compare outcomes (in both teacher competency and student learning) between groups using this framework and those using a "business-as-usual" approach.

Second, the policy and ethical dimensions will likely need continual refinement as technology and norms evolve. Legal and moral questions around generative AI (such as intellectual property rights of AI-generated images, or equity issues if not all students have equal access to AI tools) are dynamic and context-dependent. Our sample policy provides a starting template, but institutions might adapt specific clauses to their environment, and updates will be needed as, for example, new regulations come into play or AI tools change in capability (e.g., tools that leave less traceable process evidence might require new guidelines).

Third, as noted, cultural adaptability is important. Leder's model of aesthetic processing, while widely cited, was developed within a particular cultural understanding of art appreciation. Different educational systems or artistic traditions might emphasize different aspects of the process. Studies could explore applying our framework in non-Western art education contexts or in related fields like design, media arts, or music, examining what modifications are necessary. Perhaps certain stages are merged or an additional stage (like "social reflection" in community-based art) is considered.

Finally, the ultimate measure of success will be the long-term development of learners. If we implement this framework widely, do students (future artists and designers) show greater growth in creative thinking skills, critical reflection, and ethical tech use? Do they emerge as professionals who are adept at using AI as a tool without losing their unique creative voice and skills? The goal is not merely to manage AI's presence, but to elevate the quality of art education in its presence. We hypothesize that by being transparent about process and by positioning AI as a collaborator rather than an auto-pilot, students will indeed cultivate higher-order thinking and meta-cognitive skills that serve them beyond the classroom. Verifying this through longitudinal research or case studies will be an important next step.

In summary, we have introduced a novel synthesis of pedagogical and cognitive theory tailored to the challenges and opportunities of generative AI in art education. The AI-Aesthetic TPACK framework and its associated tools reposition teachers and students from being passive users of AI technology to active designers of learning experiences with AI. By ensuring that AI's role remains that of a scaffold—supporting and extending human creativity rather than replacing it—we aim to preserve the integrity of the artistic learning process. The contributions here provide both a vision and a practical starting kit for educators venturing into this new territory. We invite further collaboration and research to refine these tools, test them in diverse settings, and collectively shape an art education that is enriched, not eclipsed, by artificial intelligence.

ACKNOWLEDGEMENT

The author thanks colleagues and peer reviewers for their insightful feedback on earlier drafts of this work. No human subjects were directly involved in this study's data collection, and no external funding was received.

REFERENCES

[1]. Abdullahi Yusuf, A., Pervin, N., & Román-González, M. (2024). Generative AI and the future of higher education: A threat to academic integrity or reformation? Evidence from multicultural perspectives. International Journal of Educational Technology in Higher Education, 21(1), Article 21. https://doi.org/10.1186/s41239-024-00453-6

[2]. Agudo, I., Arrese, J., Liberal, A., & Matute, H. (2022). Assessing emotion and sensitivity of AI artwork. Frontiers in Psychology, 13, 879088. https://doi.org/10.3389/fpsyg.2022.879088

[3]. Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology: General, 144(1), 114–126. https://doi.org/10.1037/xge0000033

[4]. Horton, C. B., White, M. W., & Iyengar, S. S. (2023). Bias against AI art can enhance perceptions of human creativity. Scientific Reports, 13(1), Article 19001. https://doi.org/10.1038/s41598-023-45202-3

[5]. Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. British Journal of Psychology, 95(4), 489–508. https://doi.org/10.1348/0007126042369811

[6]. Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. Organizational Behavior and Human Decision Processes, 151, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

[7]. Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. Teachers College Record, 108(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

[8]. Natarajan, S., Mathur, S., Sidheekh, S., Stammer, W., & Kersting, K. (2025). Human-in-the-loop or AI-in-the-loop? Automate or collaborate? In Proceedings of the AAAI Conference on Artificial Intelligence, 39(27) (pp. 28594–28600). https://doi.org/10.1609/aaai.v39i27.35083

[9]. Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., et al. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. Annals of Internal Medicine, 169(7), 467–473. https://doi.org/10.7326/M18-0850

[10].Wen, W., & Wen, H. (2024). Bridging generative AI technology and teacher education: Understanding preservice teachers' processes of unit design with ChatGPT. Contemporary Issues in Technology and Teacher Education, 24(4). Retrieved from https://citejournal.org/volume-24/issue-4-24/general/bridging-generative-ai-technology-and-teacher-education-understanding-preservice-teachers-processes-of-unit-design-with-chatgpt

[11].Zhang, Y., & Gosline, R. (2023). Human favoritism, not AI aversion: People's perceptions (and bias) toward generative AI, human experts, and human–GAI collaboration in persuasive content generation. Judgment and Decision Making, 18(6), 1485–1503. https://doi.org/10.1017/jdm.2023.37