Volume 9, Issue 8, 2026 ISSN: 2617-9938
DOI: https://doi.org/10.31058/j.ad.2026.98009

Task Delegation and Student Judgement in Design Education with Generative Artificial Intelligence

Han Ni[1*], Xiaoheng Li[1]

[1] University of Shanghai for Science and Technology, Shanghai, China
*Correspondence: Han Ni

Received: 18 September 2026**; Accepted:** 9 October 2026**; Published:** 9 November 2026

Abstract:

Generative artificial intelligence increasingly participates in problem framing, information organisation, evaluation and production in design education. This study examines how task delegation relates to students' control over design judgement. A qualitative multiple-case study analysed ongoing human–AI conversations involving three design learners across four projects. Of 70 candidate episodes, 45 were included following an eligibility review. Researcher-led coding distinguished design phase, student agency moves, AI role, review behaviour, delegation category, interaction trajectory and judgement ownership. The supervisor proposed candidate codes, which the researcher checked against relevant excerpts from the source conversations before making the final coding decisions. Selected and borderline cases were discussed with the supervisor. Delegation and judgement ownership showed heterogeneous configurations. Among 13 production-delegation episodes, eight were human-dominant, three human-led shared, one AI-led shared and one AI-dominant. Three of ten structured-support episodes were AI-led shared. Explicit review was observed in 11 episodes, all classified as human-dominant or human-led shared; however, 22 of the 34 episodes without explicit review also received these classifications. These descriptive findings distinguish task completion from control over problem boundaries, criteria and direction. We propose five pedagogical checkpoints addressing framing, verification, divergence, criteria and accountability. The checkpoints are interpretive proposals rather than validated interventions. The study suggests that assessment of AI-supported design work should examine students' reasons for accepting, rejecting and revising contributions alongside the artefacts produced. The small, unevenly distributed corpus and interpretive coding limit generalisation and do not establish effects on learning or creativity.

Keywords:

Generative Artificial Intelligence, Design Education, Human–AI Co-creation, Design Judgement, Task Delegation, Judgement Ownership, Ai Literacy, Review Behaviour

1. Introduction

Generative artificial intelligence (GenAI) is becoming part of students' everyday design work, including research, concept development, visual generation, technical troubleshooting and written communication. Fleischmann's survey-based study

argues for explicit integration of GenAI into design curricula, including critical understanding of its capabilities and limitations [1]. Ma et al. examine how generative feedback enters students' judgement practices through recurring design iterations [2]. These developments require educators to consider how responsibility for design decisions is distributed during tool use.

Design learning involves more than producing a finished artefact. Students learn through framing uncertain problems, evaluating competing proposals and revising their intentions in response to materials and feedback. Schön's account of reflection-in-action and Cross's account of designerly knowing provide a basis for examining these processes [3, 4]. When production becomes easier to delegate, the resulting artefact may provide insufficient evidence of who established its purpose or assessed its adequacy.

Research outside design education identifies related concerns. Fan et al.'s randomised writing study distinguishes assisted task performance from knowledge gains and learning processes [5]. Lee et al.'s survey of 319 knowledge workers links greater confidence in GenAI with less self-reported critical thinking and identifies verification, integration and task stewardship as important forms of thinking during AI use [6]. These findings concern particular tasks and self-reports; they do not establish that GenAI necessarily impairs design learning. A scoping review similarly distinguishes uses that support learner agency from uses associated with dependence and cognitive offloading [7].

Within design education, Naik et al. analyse students' reflections on judgements concerning reliability, appropriateness, output quality and agency distribution [8]. Their work raises a process-level question: how can researchers identify the retention or transfer of judgement in the interaction itself? Studies of collaborative prompting and workplace innovation further indicate that collaboration arrangements and outcomes require separate examination [9, 10]. Task delegation, reliance and control over judgement should therefore not be treated as interchangeable measures.

This study addresses three research questions. RQ1: How are AI roles, delegation categories and judgement ownership configured across design phases and task contexts? RQ2: Can the same delegation category coexist with different configurations of judgement ownership? RQ3: What review behaviours and interaction trajectories are visible in these configurations, and what pedagogical checkpoints might they inform?

The contribution is an episode-level analysis of naturally occurring design conversations that separates responsibility for task execution from control over consequential design judgements. The analysis combines descriptive distributions with contrasting episodes. It then develops pedagogical proposals for making students' judgement more visible. It does not measure learning gains or test the effectiveness of the proposed checkpoints.

2. Related Research and Conceptual Framework

2.1. Generative feedback and design learning

The integration of GenAI into design curricula involves decisions about when to use it, how to evaluate its contributions and which responsibilities remain with learners [1]. Ma et al. describe AI-assisted divergent pedagogy as a process-oriented interpretation of generative feedback in design thinking education [2]. Together with

Naik et al.'s account of student judgement [8], this work directs attention towards how generated material is considered, modified and accepted within a learning activity.

Creativity research also cautions against equating easier production with broader exploration. In a short-story experiment, Doshi and Hauser found that AI assistance could improve individual creative evaluations while reducing diversity across outputs [11]. This finding motivates attention to alternative generation, but its transfer to design tasks remains a question rather than an established result. The present study does not evaluate the originality or diversity of final designs.

Design pedagogy offers reasons to preserve opportunities for deliberation. Rutherford discusses desirable difficulty in relation to durable design learning [12], while Laamanen et al. describe material-led processes involving conceptual inquiry, challenge and identity formation [13]. Wadinambiarachchi et al. investigate multimodal prompting as a means of supporting reflection during AI-assisted ideation [14]. These studies inform the proposition that some effortful evaluation may have educational value even when routine production can be accelerated.

2.2. Delegation and judgement ownership

Design judgement concerns what matters in a situation, which criteria are appropriate and when a proposal is acceptable. Such judgements often depend on incomplete information and competing values. Reflection and critique help make their grounds communicable [3,15]. Research on peer critique also treats evaluative exchange as part of design learning rather than merely commentary on finished work [16].

We use judgement ownership as an operational term for visible control over five activities within an episode: framing the problem, setting consequential constraints, establishing criteria, evaluating an output and deciding the direction of further work. Ownership here does not denote legal ownership, intellectual property or a stable personal attribute. AI-led classifications refer to the apparent influence of generated contributions on the recorded direction of work, not to moral responsibility or consciousness attributed to a system.

Delegation concerns the task responsibility assigned to AI. A student may delegate image production while specifying the intended scene and acceptance criteria. Conversely, a student may request only a project outline but adopt an AI-generated organisation of the problem. We therefore code delegation and judgement ownership separately. The D0–D6 labels distinguish categories of delegated responsibility; they are not equal intervals on a validated scale of cognitive effort or agency.

2.3. Review behaviour and pedagogical scaffolding

Observable review includes checking claims, challenging a proposal, rejecting an output, articulating criteria and revising constraints in response to AI contributions. Buçinca et al. show that cognitive forcing interventions can reduce overreliance in an AI-assisted decision task, although effectiveness and subjective experience involve trade-offs [17]. Hou et al. examine a critical-thinking intervention in problem-based human–GenAI learning [18]. These studies support investigating structured evaluation opportunities without establishing that any particular checkpoint will work in a design studio.

The OECD and European Commission framework organises AI literacy around engaging with, creating with, managing and shaping AI [19]. Its intended setting is primary and secondary education. We draw on its emphasis on evaluation and responsibility as a conceptual resource, rather than claiming that it validates a higher-education assessment instrument. Vendrell and Johnston likewise propose scaffolding critical thinking during large-language-model use [20]. In this study, review capability denotes situated practices that make judgement inspectable; it is not a psychometric construct or a score inferred from prompt frequency.

3. Materials and Methods

3.1. Research design and corpus

This qualitative multiple-case study analyses course-related ChatGPT conversations produced by three postgraduate design learners at the University of Shanghai for Science and Technology, including one author of this article. The three learners contributed four projects involving service design, interaction design and technical work in Rhino, a three-dimensional modelling application. The author provided the FC01 Sam's Club service-design conversations and also participated in data analysis; P02 contributed both FC02 and FC03. Conversations took place between early March and early August 2026 as part of learners' design-course activities; the researcher did not intervene in the other learners' original exchanges. The conversations were conducted using ChatGPT. Response-level model identifiers could not be verified from the available records; therefore, no specific model version is attributed to the analyzed responses. The analysis compares episode configurations and does not estimate change over time.

Records were organised at four levels: project, conversation, episode and exchange. A project is a continuing design undertaking; a conversation is a continuous chat; an episode groups exchanges around a relatively identifiable design objective; an exchange is one user contribution and the corresponding AI response. The episode is the unit of coding. Multiple episodes from the same person and project are not independent participants or independent observations for population inference.

Three learners supplied the conversations analysed in this study; the author's own Sam's Club conversations formed one of these cases. Seventy candidate episodes across their four projects were assessed for relevance to a focal design problem, sufficient visible context and evidence of delegation, evaluation, selection, revision or acceptance. Forty-five were included, 22 were excluded for insufficient context and three were retained as contextual material outside formal analyses. These are episode counts, not a count of additional participants or projects. Table 1 reports the included corpus; Figure 1 summarises the workflow. P02 contributed both FC02 and FC03. The main comparison between FC01 and FC03 involves different learners as well as different tasks, so task and participant effects cannot be separated.

Table 1. *Project composition of the included corpus.*

Project	Learner	Project context	Main task context	n
FC01	P01	Sam's Club service design	Service system and scenario design	26
FC02	P02	OPC interaction design	Trust, oversight and information structure	3
FC03	P02	Rhino 3D modelling	Technical execution and geometric troubleshooting	7
FC04	P03	Decathlon service design	Outdoor product–service systems	9

Project and participant identifiers are used throughout the analysis. Project descriptions identify task contexts rather than participants. FC02 and FC03 involve the same learner; FC01 and FC03 do not.

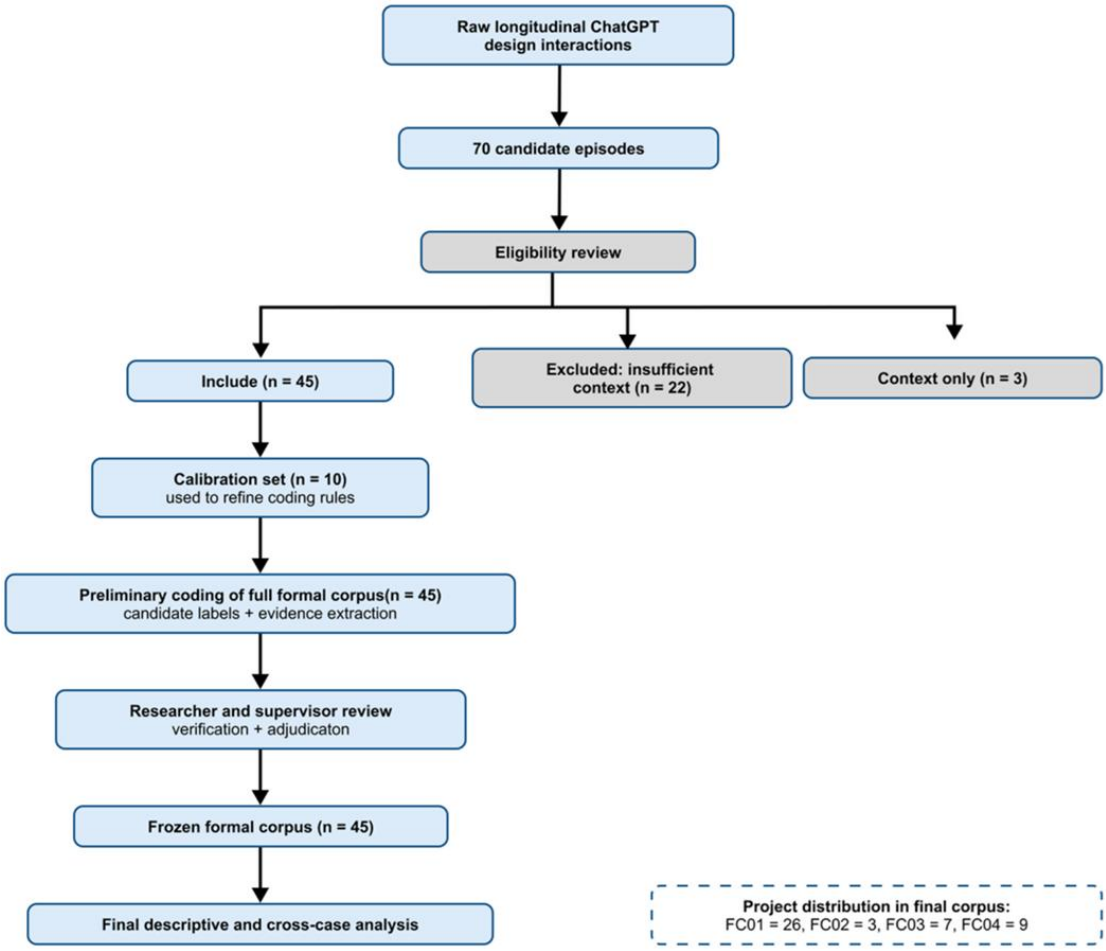


Figure 1. *Corpus selection and coding workflow. The calibration set was drawn from the included corpus and is not an additional sample. Context-only material is excluded from all totals based on N = 45.*

3.2. Coding dimensions and decision rules

The coding framework distinguishes seven dimensions. Design phase identifies the primary activity as framing, research, ideation, evaluation, development or execution. Student agency moves identify actions such as initiating, constraining, reframing, comparing, selecting, rejecting, verifying, defining criteria, giving reasons and delegating. AI role identifies information provision, tutoring, ideation partnership, critique, recommendation, decision support, execution or editing. AI roles and review behaviours may receive multiple labels within an episode.

The delegation categories are D0 self-led, D1 information support, D2 structured support, D3 recommendation, D4 decision delegation, D5 production delegation and D6 final-output delegation. The labels preserve the coding scheme while distinguishing qualitatively different responsibilities. In particular, production delegation and decision delegation should not be interpreted as successive quantities of the same cognitive task. No means, correlations or tests of a linear trend across D0–D6 are calculated.

Judgement ownership is coded as human-dominant, human-led shared, negotiated, AI-led shared, AI-dominant or unclear. The assessment considers visible problem definition, constraints, criteria, evaluation and directional choice. The categories describe control as evidenced within the episode. They cannot establish what the learner thought privately or decided outside the recorded interaction.

Interaction trajectory describes the overall progression as linear acceptance, iterative refinement, critical negotiation, divergent exploration, human-led convergence, AI-led completion or unclear. AI-led completion denotes substantial AI-driven progression towards completion; it does not, by itself, determine judgement ownership. Repeated prompting is not sufficient evidence of critical negotiation.

Review behaviour is distinguished from an initial instruction. A request to change language or move an image element is a constraint unless the surrounding sequence shows evaluation of an identifiable AI contribution. Explicit review requires visible checking, evaluation, challenge, correction or rejection after an AI contribution. The label no review observed means that such behaviour was not visible within the episode; it does not mean that the student made no judgement. A quoted sentence alone is therefore illustrative rather than sufficient to reproduce the whole-episode classification.

3.3. Calibration and researcher verification

The analysis used ten calibration episodes selected for variation across the four projects, conceptual and execution tasks, delegation categories and review behaviour. The researcher and supervisor examined typical and borderline cases to refine coding rules. These rules included separating high production delegation from AI-led judgement, distinguishing repeated prompt editing from review, and assessing trajectories from the visible sequence rather than response length.

The supervisor proposed candidate codes. The researcher located relevant excerpts, checked the proposed codes against the source conversations and made the final coding decisions. Selected and borderline cases were discussed with the supervisor; this process did not constitute independent double coding.

Partial codability was allowed. When missing material prevented interpretation of a trajectory but still allowed delegation and ownership to be assessed, trajectory was marked unclear rather than inferred. Three episodes lacked sufficiently stable primary-phase descriptions, and two lacked sufficient evidence for trajectory classification. Missing classifications are identified in the relevant distributions rather than treated as substantive design phases or trajectories.

3.4. Descriptive analysis and interpretive boundaries

Counts and cross-tabulations describe the included corpus. Unless otherwise stated, percentages use all 45 included episodes as their denominator; primary-phase counts sum to 42 because three episodes were not assigned a stable phase. Within-project percentages use 26 episodes for FC01 and seven for FC03. Counts accompany comparative percentages where small denominators are particularly consequential. Multi-label counts must not be summed as if they represented distinct episodes.

Contrasting episodes were used to clarify how similar delegation categories could coexist with different ownership configurations. The pedagogical checkpoints were subsequently developed through interpretation of these patterns alongside the

literature. They were not an intervention experienced by participants, and the corpus does not test their effects. Review and ownership were coded from overlapping evidence in the same conversations, so their co-occurrence is not an independent validation of either construct.

3.5. Ethics and representation

Participants supplied exported course-related conversations for research use. The author supplied the FC01 conversations; the two other learners supplied their own records. All records were checked by the researcher and discussed with the supervisor; this was not independent verification. Participant identifiers P01–P03 and project identifiers FC01–FC04 are used in the analysis. Short excerpts relevant to the reported judgements are reproduced. The English excerpts are translations of Chinese-language speech; summaries of AI responses are paraphrases. The analysis does not infer stable abilities or psychological states from individual exchanges.

The two contributing classmates were informed that their anonymized conversations would be used in a paper and quoted publicly; signed permissions were subsequently obtained and retained by the researcher. The author contributed their own conversations. The researcher keeps the source material on a personal computer and has prepared anonymized copies for sharing.

4. Results

4.1. Design phases and AI roles

Of the 45 episodes, 42 received a stable primary design-phase code. Execution accounted for 14 episodes (31.1% of 45); framing and development each accounted for ten (22.2%); research accounted for six (13.3%); ideation and evaluation each accounted for one (2.2%). Three episodes (6.7%) were not assigned a stable phase. The corpus therefore includes substantial framing and execution work, but provides limited coverage of ideation and evaluation as primary activities.

AI roles were multi-label: decision support occurred in 15 episodes, execution in 13, tutoring in six, editing and information provision in five each, ideation partnership and recommendation in four each, and critique in one. Table 2 reports selected prominent roles by phase, rather than an exhaustive role-by-phase cross-tabulation. Eight of ten framing episodes involved D3 recommendation, while nine of 14 execution episodes involved D5 production delegation. Research episodes were distributed across information support, structured support, recommendation and final-output delegation.

Table 2. *Primary design phase, selected AI roles and delegation categories.*

Design phase	n	Selected AI roles (counts)	Delegation categories
Framing	10	Decision support (6); recommendation (4)	D2 = 2; D3 = 8
Research	6	Information source (4); decision support (2)	D1 = 3; D2 = 1; D3 = 1; D6 = 1
Ideation	1	Ideation partner (1)	D3 = 1
Evaluation	1	Critic (1)	D2 = 1
Development	10	Decision support (6); executor (3); editor (2)	D1 = 1; D2 = 3; D3 = 3; D5 = 3
Execution	14	Executor (9); tutor (4); editor (2)	D0 = 1; D1 = 1; D2 = 3; D5 = 9
Unassigned	3	Not tabulated	D3 = 2; D5 = 1

The role column lists selected prominent roles only and is not intended to reproduce the full role totals. AI roles are multi-label. Delegation counts include all 45 episodes; the unassigned row refers only to primary phase.

4.2. Delegation categories and judgement ownership

D3 recommendation was the most frequent delegation category (15/45, 33.3%), followed by D5 production delegation (13/45, 28.9%), D2 structured support (10/45, 22.2%) and D1 information support (5/45, 11.1%). D0 self-led and D6 final-output delegation each occurred once (2.2%). No D4 decision-delegation episode was recorded. Ownership was human-led shared in 23 episodes (51.1%), human-dominant in ten (22.2%), AI-led shared in ten (22.2%) and AI-dominant in two (4.4%). No episode was classified as negotiated or unclear for ownership.

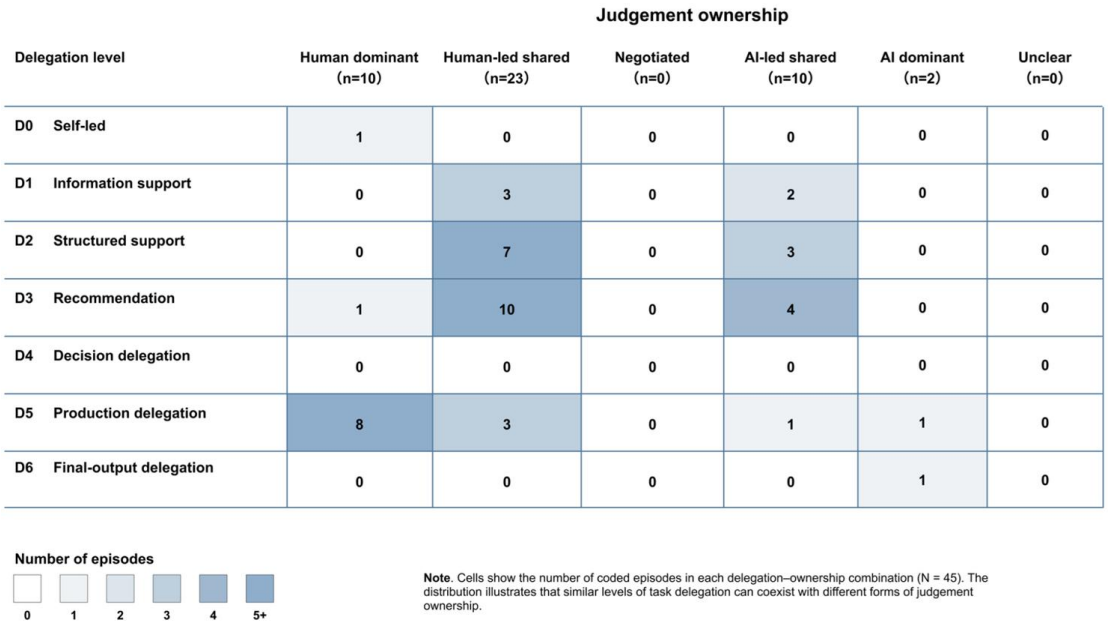


Figure 2. *Cross-tabulation of delegation category and judgement ownership for 45 episodes. Cell values are episode counts. Category numbers do not define an interval scale, and the figure does not estimate a statistical trend.*

The production-delegation category contained all four observed ownership configurations: eight human-dominant, three human-led shared, one AI-led shared and one AI-dominant. In SAM-E07, the learner delegated image production but specified an intended visual character: "I want this image to match the style of the second image: a realistic scene, not an illustration." The episode was coded D5, human-dominant and iterative refinement. In contrast, SAM-E12 began with a request to generate a storyboard from a user journey; the AI selected the focal scenes, characters, narrative tension and eight-frame structure. It was coded D5, AI-dominant and AI-led completion. These are contrasting configurations within the same delegation category, not measures of how much the learner learned in either episode.

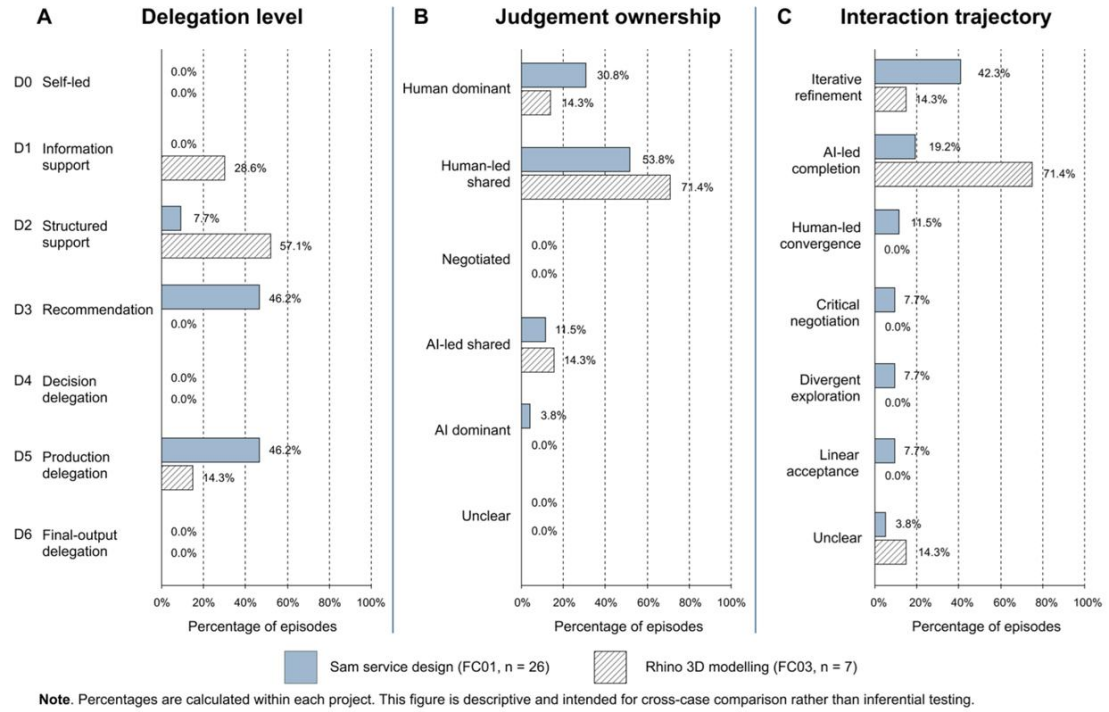
AI-led shared ownership also occurred in three of the ten D2 episodes; the remaining seven were human-led shared. D3 contained ten human-led shared, four AI-led shared and one human-dominant episode. DEC-E04 illustrates the conceptual relevance of structured support. The learner asked what work should be undertaken across the early, middle and later stages of a smart product–service system project. The AI organised the visible project structure and proposed the path forward. The

episode was coded D2 and AI-led shared even though the request did not delegate production of the final design.

The cross-tabulation supports a bounded conclusion: delegation category does not uniquely specify judgement ownership in this corpus. It does not establish a non-linear statistical relationship or a general ordering of categories by educational risk. The nature of the delegated responsibility and the available evidence of student evaluation both matter to interpretation.

4.3. Service design and technical modelling

FC01 service design and FC03 Rhino modelling offer contrasting task contexts (Figure 3). In FC01, recommendation and production delegation each occurred in 12 of 26 episodes (46.2%), with structured support in two (7.7%). FC03 contained four structured-support episodes (57.1%), two information-support episodes (28.6%) and one production-delegation episode (14.3%). These distributions describe the two projects; the comparison does not isolate an effect of task type.



Note. Percentages are calculated within each project. This figure is descriptive and intended for cross-case comparison rather than inferential testing.

Figure 3. *Within-project distributions for FC01 service design (n = 26) and FC03 Rhino modelling (n = 7). Panels show delegation, ownership and trajectory. Percentages are descriptive; projects involve different participants and unequal numbers of episodes. Unclear trajectory denotes insufficient evidence.*

Ownership in FC01 was human-led shared in 14 episodes (53.8%), human-dominant in eight (30.8%), AI-led shared in three (11.5%) and AI-dominant in one (3.8%). In FC03, five episodes were human-led shared (71.4%), one human-dominant and one AI-led shared (14.3% each). AI-led completion occurred in five FC03 episodes (71.4%), compared with five FC01 episodes (19.2%). Progression towards completion and judgement ownership thus produced different descriptions of the same project contexts.

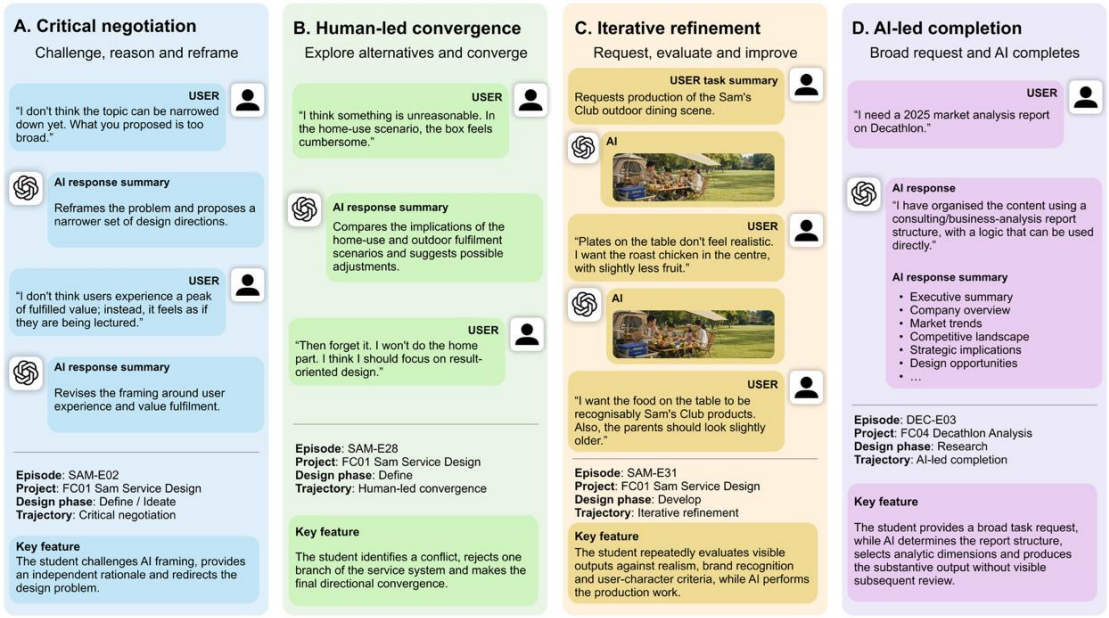
RHI-E02 illustrates how external feedback could enter the interaction. After receiving Rhino advice, the learner reported continuing difficulties and redirected the task: “There still seems to be a problem. Let's try another approach: instead of making

a surface, let's reconsider how to close the curve." The episode was coded D2, human-led shared and iterative refinement, with verification and challenge. The visible exchange suggests that the learner used the reported state of the model to reassess the suggested approach. It does not independently establish that the eventual geometry was correct.

In service design, recommendations may influence the organisation of a problem even without producing a final artefact. In modelling, detailed procedural assistance may coexist with student evaluation of software feedback. These examples motivate distinguishing the target of delegation from the amount of generated material.

4.4. Interaction trajectories

AI-led completion occurred in 17 episodes (37.8%), iterative refinement in 13 (28.9%), divergent exploration and human-led convergence in four each (8.9%), linear acceptance in three (6.7%) and critical negotiation in two (4.4%). Two episodes (4.4%) were marked unclear because the available multimodal record was insufficient to reconstruct their trajectories. (Figure 4)



Note. Representative episodes illustrate four observable human–AI interaction trajectories. User quotations were translated from Chinese and shortened where necessary. Boxes labelled "AI response summary" or "USER task summary" are paraphrases rather than verbatim quotations.

Figure 4. *Four illustrative interaction trajectories. The panels summarise recorded episode interpretations; translated learner excerpts are marked with quotation marks. AI contributions are paraphrased. These examples are not an exhaustive taxonomy or independent evidence of learning outcomes.*

Of the 17 AI-led completion episodes, eight were human-led shared, seven AI-led shared and two AI-dominant. All 13 iterative-refinement episodes were either human-led shared (seven) or human-dominant (six). Both critical-negotiation episodes were human-led shared. The four human-led convergence episodes comprised two human-dominant and two human-led shared cases. These cross-tabulations reinforce the need to distinguish progression from evaluative control.

In SAM-E02, critical negotiation was visible when the learner challenged premature closure and an overly broad framing, then questioned whether the proposed experience delivered a meaningful benefit. In SAM-E28, human-led convergence involved rejecting the home-use branch because the container appeared cumbersome

and deciding to remove that branch. Iterative refinement describes repeated adjustment within an established direction, whereas AI-led completion describes substantial AI progression towards an output. Ownership must still be assessed separately in each trajectory.

4.5. Explicit review behaviour

Explicit review was observed in 11 of 45 episodes (24.4%). The remaining 34 (75.6%) were coded no review observed. Multi-label counts were ten for modifying constraints, seven for rejecting outputs, six for articulating criteria, five for challenging AI, four for correcting AI and one for verifying a claim. Table 3 links these counts to short illustrative excerpts; the relevant sequence, rather than the excerpt alone, supplied the basis for coding.

Table 3. *Review behaviours and illustrative translated excerpts.*

Behaviour	n	Episode	Translated excerpt
Challenge AI	5	SAM-E02	"I do not think the topic can be narrowed down yet. What you proposed is too broad."
Articulate criteria	6	SAM-E07	"I want this image to match the style of the second image: a realistic scene, not an illustration."
Reject output	7	SAM-E28	"I think some aspects are unreasonable. In the home-use scenario, the box feels cumbersome."
Verify claim	1	RHI-E02	"There still seems to be a problem. Let's try another approach: instead of making a surface, let's reconsider how to close the curve."
Modify constraints	10	SAM-E24	"I do not think my product should instruct users how to behave. Compartmentalising the shopping basket may affect their shopping experience."
Correct AI	4	SAM-E07	"…a realistic scene, not an illustration."
No review observed	34	SAM-E43	"I need it entirely in English; keep only the product images and the basic process." (Contrast case without explicit review.)

Counts refer to episodes bearing each label. Labels may overlap. The same episode may illustrate more than one behaviour. No review observed is an absence-of-observed-behaviour code, not evidence of absent judgement. All excerpts are translated from Chinese.

The 11 episodes with explicit review comprised six human-led shared and five human-dominant episodes. However, 22 of the 34 episodes without explicit review were also human-led shared (17) or human-dominant (five). Explicit review therefore co-occurred with human-led classifications, but its absence did not imply AI-led ownership. Pre-existing goals or constraints could support an ownership interpretation without a visible challenge to an output.

This pattern must be interpreted cautiously. Review and ownership share some evidential features, such as the articulation of criteria and visible directional choice. Their co-occurrence may partly reflect the coding definitions and researcher interpretation. It cannot establish that review caused students to retain judgement, nor validate review counts as measures of critical-thinking ability.

5. Proposed Pedagogical Checkpoints

We propose five checkpoints for making judgement visible during design work: framing before generation, verification before reliance, divergence before convergence, criteria before choice and accountability before submission (Figure 5). They are design propositions informed by the corpus and literature. Framing, verification and criteria have direct illustrative episodes in the analysis; divergence and accountability extend the observations through pedagogical reasoning. None has been tested as an intervention in this study.

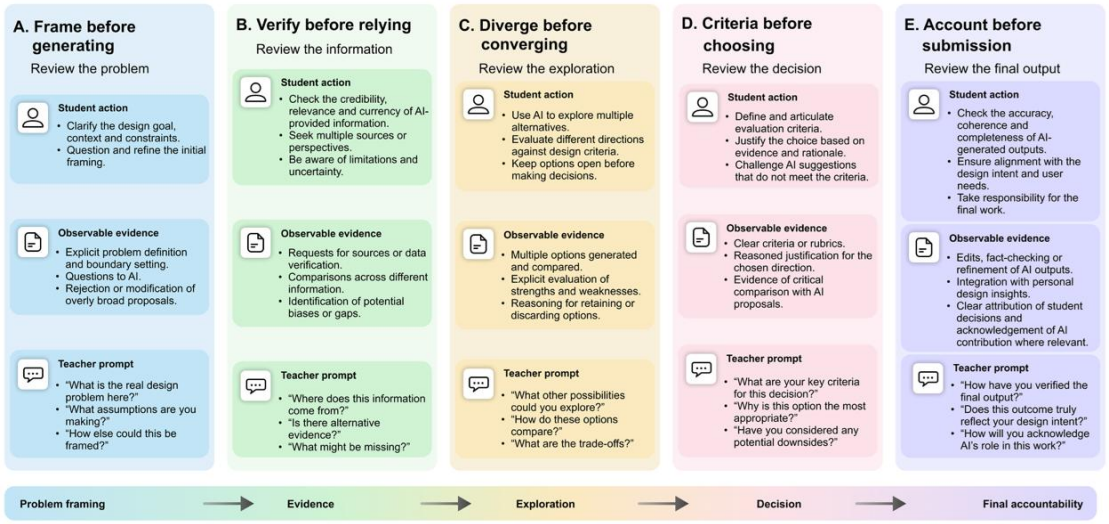


Figure 5. *Proposed checkpoints for AI-supported design education. Each checkpoint links a design activity to evidence a learner could provide. The sequence is indicative: checkpoints may recur as a project develops. The framework has not been experimentally validated.*

5.1. Framing before generation

Before requesting broad AI-generated directions, students could record an initial problem statement, intended users, important assumptions and project boundaries. They could then compare the AI's proposals with that initial position and explain any revisions. The purpose is to make changes in framing visible, not to prevent useful AI suggestions.

SAM-E02, SAM-E24 and SAM-E28 provide examples of challenging or revising a direction. In SAM-E24, the learner argued that a product should not instruct users how to behave and that compartmentalising a shopping basket might impair the shopping experience. A teacher could ask which assumption was challenged and why the revised boundary better served the intended experience. Such evidence allows a framing decision to be examined without inferring learning from the finished image alone.

5.2. Verification before reliance

For consequential factual or technical claims, students could record the AI contribution, the external source or test used to check it, and whether checking changed the decision. RHI-E02 illustrates procedural checking through reported software feedback. This checkpoint distinguishes a generated suggestion from the

evidence used to accept it. Verification requirements should be proportionate to the consequences of an error and the learning objective.

5.3. Divergence before convergence

Before selecting a proposal, students could compare alternatives grounded in genuinely different assumptions, user priorities or resource constraints. Asking for more outputs alone would not demonstrate this difference. Doshi and Hauser's findings about convergence across AI-assisted stories motivate attention to diversity [11], but do not establish the effect of this proposed design checkpoint. Students might explain how alternatives differ in their assumptions and what would be lost by choosing each one.

5.4. Criteria before choice

Students could state provisional criteria before asking AI to compare options. These might concern feasibility, accessibility, experience or burdens on other participants in a service. AI could then critique the criteria or apply them provisionally, while the student explains the eventual selection. This proposal is consistent with work on articulating design criticism [15] and structured critical-thinking support [18]. Criteria may be revised during exploration; the relevant evidence is an explanation of the revision, not an immutable initial checklist.

5.5. Accountability before submission

At submission, students could provide a brief account of delegated tasks, retained decisions, consequential revisions and one important AI contribution that they accepted or rejected. This would complement tool disclosure with reasons for endorsing the final design. It is a pedagogical proposal about answerability for decisions, not a legal test of authorship. Its usefulness and workload would need to be examined in teaching practice.

6. Discussion

6.1. Separating production from evaluative control

The most consequential result is the diversity of ownership configurations within production delegation. The same category included both human-dominant and AI-dominant episodes. Conversely, structured support sometimes involved AI-led shared judgement. A count of generated outputs or requests would not capture these distinctions. This interpretation is consistent with broader evidence that collaboration arrangements and innovation outcomes require separate analysis [10].

Han et al. examine collaboration strategies during team prompting [9]. The present analysis focuses on episode-level control in individual learners' conversations. It does not establish a superior measure of co-creation; it offers a complementary description of who visibly framed, evaluated and redirected particular tasks. Any instructional use should retain this situational meaning.

6.2. Evidence of learning in AI-supported design work

Sketches, models and prototypes remain important sites of design reasoning. The present findings do not justify replacing making with judgement statements. They suggest supplementing artefacts with evidence of problem framing, criteria and

consequential revisions when production is delegated. Reflection-in-action and critique remain useful conceptual resources for this extension [3, 12, 14].

The Rhino comparison illustrates why technical progress alone is ambiguous. AI-led completion was frequent even when ownership was human-led shared. Conversely, support for organising a project could influence direction without producing the final artefact. Educators might therefore examine the nature of the responsibility delegated and the learner's grounds for endorsement rather than assign a single permissible percentage of AI use. This is an implication for investigation, not an assessment rule validated by the present data.

6.3. Review as situated evidence

Review behaviours make some reasoning available for inspection, but they are incomplete traces of judgement. A student may exercise control through an initial specification, external work or a decision not recorded in chat. The 22 human-led episodes without explicit review demonstrate why absence of an observed challenge cannot be treated as absence of agency.

Lee et al.'s account of verification, integration and task stewardship helps contextualise the review behaviours [6]. Cognitive forcing research also indicates that demands for deliberation can impose interaction costs [17]. Pedagogical checkpoints should therefore be selective and responsive to the task. Requiring a written reflection after every prompt could create administrative work without providing better evidence of learning.

6.4. Implications for curriculum and assessment

A process portfolio could include the initial problem frame, criteria revisions, selected rejected outputs and reasons for key decisions. During critique, teachers could ask learners to explain a consequential acceptance or rejection using the project's constraints and external evidence. These materials would complement, rather than replace, assessment of the artefact and the student's ability to make or modify it.

Different activities warrant different questions. Framing calls for attention to boundaries and assumptions; research calls for source verification; exploration calls for comparison; evaluation calls for defensible criteria; execution calls for clear acceptance conditions and checking. The limited representation of primary ideation and evaluation episodes means that recommendations for those phases rely particularly on literature-informed interpretation.

These proposals situate evaluation and responsibility within AI literacy [19]. They also align with calls for deliberate curricular integration [1] and critical-thinking scaffolds [20]. They should be adapted to course objectives and tested for usefulness, fairness and workload before being adopted as assessment standards.

6.5. Limitations and further research

The corpus contains three learners and four unevenly represented projects. Twenty-six episodes come from the author’s FC01 case and only three from FC02, so the distributions are sensitive to case selection and project activity. The author acted both as a participant supplying conversations and as an analyst; familiarity with the course context may have aided interpretation but could also have influenced selection and coding. Episode counts are not participant sample sizes. The comparison of service

design and Rhino cannot separate participant differences from task differences, and it does not establish a developmental trajectory over time.

Missing multimodal outputs limited some interpretations, and the conversations do not capture all work undertaken outside the chat. The analysis relied on the researcher's judgement: the supervisor proposed candidate codes, while the researcher located relevant excerpts, checked the proposed codes against the source conversations and made the final coding decisions. Selected and borderline cases were discussed with the supervisor, but this process did not constitute independent double coding. Shared evidence between review and ownership may amplify their apparent association. Future work could use an independently coded subset, clearer disagreement records and access to relevant artefacts to assess interpretive robustness.

The study does not measure creativity, critical-thinking ability, retention or transfer. It also does not test the five checkpoints. Subsequent studies could compare supported and unsupported AI use within comparable course activities, examine the quality of students' reasons and resulting designs, and assess whether any benefits persist beyond the immediate task. Such work would need to account for participant experience, model changes and the burden of documentation.

7. Conclusions

In this corpus, task delegation and judgement ownership provide distinct descriptions of AI-supported design activity. Production could be delegated while students retained visible control over criteria and direction, and structured support could influence the framing of a project. Explicit review offered useful evidence of some decisions, but its absence did not establish a loss of judgement. These distinctions support examining the reasons behind accepted, rejected and revised contributions alongside the resulting artefacts.

The five proposed checkpoints translate this interpretation into questions that can be explored in studio teaching. Their value remains to be tested. The educational issue is how to make students' consequential judgements inspectable and answerable when parts of the design process are delegated, while retaining opportunities to learn through making.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Author Contributions

H.N., Conceptualization, Designing the study question, Statistical analysis, Methodology, Validation, Writing—original draft, Project administration.

X.L., Investigation, Acquisition of data, tables, Statistical analysis, Validation, Revision of the manuscript.

All authors have read and agreed to the published version of the manuscript.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

AI Assistance in Manuscript Preparation

During manuscript preparation, AI tools were used for language and formatting assistance to enhance readability and language, and to check and correct errors or biases in expression. The authors reviewed the resulting text and remain responsible for the manuscript.

Informed Consent Statement

Informed consent was obtained from all subjects involved in the study.

Data Availability Statement

Upon reasonable request, data used and analyzed during the current study are available from the corresponding author.

References

[1] Fleischmann, K. Making the case for introducing generative artificial intelligence (AI) into design curricula. *Art Des. Commun. High. Educ.* 2024, 23(2), 187-207, DOI: 10.1386/adch_00088_1.

[2] Ma, Y.; Cai, Y.; Chen, J. AI-assisted divergent pedagogy: Understanding human-AI co-creativity in design thinking education. *Des. Stud.* 2026, 105, 101407, DOI: 10.1016/j.destud.2026.101407.

[3] Schön, D.A. *The Reflective Practitioner: How Professionals Think in Action*; Basic Books: New York, NY, USA, 1983; ISBN 978-0-465-06874-6.

[4] Cross, N. *Designerly Ways of Knowing*; Springer: London, UK, 2006; ISBN 978-1-84628-300-0, DOI: 10.1007/1-84628-301-9.

[5] Fan, Y.; Tang, L.; Le, H.; Shen, K.; Tan, S.; Zhao, Y.; Shen, Y.; Li, X.; Gašević, D. Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *Br. J. Educ. Technol.* 2025, 56(2), 489-530, DOI: 10.1111/bjet.13544.

[6] Lee, H.P.; Sarkar, A.; Tankelevitch, L.; Drosos, I.; Rintel, S.; Banks, R.; Wilson, N. The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, Yokohama, Japan, 26 April–1 May 2025; Association for Computing Machinery: New York, NY, USA; Article 1121, pp. 1-22, DOI: 10.1145/3706598.3713778.

[7] Wang, G.; Wang, W.; Yang, D.; Ren, J. Generative AI, Cognitive Offloading, and Learner Agency in Higher Education: A Scoping Review. *Behav. Sci.* 2026, 16(7), 1150, DOI: 10.3390/bs16071150.

[8] Naik, S.; Shukla, P.; Obi, I.; Backus, J.; Rasche, N.; Parsons, P. Tracing the Invisible: Understanding Students' Judgment in AI-Supported Design Work. In *Proceedings of the 2025 Conference on Creativity and Cognition*, Virtual, UK, 23–25 June 2025; Association for Computing Machinery: New York, NY, USA; pp. 438-442, DOI: 10.1145/3698061.3734399.

[9] Han, Y.; Qiu, Z.; Cheng, J.; LC, R. When Teams Embrace AI: Human Collaboration Strategies in Generative Prompting in a Creative Design Task. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*,

Honolulu, HI, USA, 11–16 May 2024; Association for Computing Machinery: New York, NY, USA; Article 176, pp. 1-14, DOI: 10.1145/3613904.3642133.

[10] Mascareño, J.; Wörtler, B.; Przegalińska, A.; Ciechanowski, L. When proximal collaboration with AI hinders innovation: The moderating role of idea originality and reliance on AI. *Comput. Hum. Behav. Rep.* 2026, 22, 101054, DOI: 10.1016/j.chbr.2026.101054.

[11] Doshi, A.R.; Hauser, O.P. Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Sci. Adv.* 2024, 10(28), eadn5290, DOI: 10.1126/sciadv.adn5290.

[12] Rutherford, S. Using desirable difficulty concepts to enhance durable learning in design education. *Art Des. Commun. High. Educ.* 2020, 19(1), 65–79, DOI: 10.1386/adch_00014_1.

[13] Laamanen, T.K.; Mäkelä, M.; Seitamaa-Hakkarainen, P. Material-led creative processes in the context of design education. *Art Des. Commun. High. Educ.* 2023, 22(1), 33-54, DOI: 10.1386/adch_00061_1.

[14] Wadinambiarachchi, S.; Waycott, J.; Wadley, G. Reviving Reflection-in-Action: Instilling Designerly Thinking in AI-Supported Ideation through Multimodal Prompting. In *Proceedings of the 2026 Conference on Creativity and Cognition*, London, UK, 13–16 July 2026; Association for Computing Machinery: New York, NY, USA; pp. 359-376, DOI: 10.1145/3803784.3807524.

[15] Thiessen, M. I don't know, I just like it: Exploring how design students think about criticism. *Art Des. Commun. High. Educ.* 2017, 16(2), 145-156, DOI: 10.1386/adch.16.2.145_1.

[16] Yilmaz, O.; Gelmez, K. 'How can I say this politely?': Remodelling peer critiquing in design education and revealing compassionate critiques. *Art Des. Commun. High. Educ.* 2023, 22(2), 277-296, DOI: 10.1386/adch_00069_1.

[17] Buçinca, Z.; Malaya, M.B.; Gajos, K.Z. To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-Assisted Decision-Making. *Proc. ACM Hum.-Comput. Interact.* 2021, 5(CSCW1), Article 188, 1-21, DOI: 10.1145/3449287.

[18] Hou, C.; Zhu, G.; Liu, Y.; Sudarshan, V.; Chong, J.L.L.; Zhang, F.Y.; Tan, M.Y.H.; Ong, Y.S. The effects of critical thinking intervention on reliance behaviors, problem-solving quality, and creativity during human-Generative AI collaborative learning. *Comput. Educ.* 2026, 247, 105576, DOI: 10.1016/j.compedu.2026.105576.

[19] OECD; European Commission. *Empowering Learners for the Age of AI: An AI Literacy Framework for Primary and Secondary Education*; OECD Publishing: Paris, France, 2026, ISBN 978-92-68-40754-7, DOI: 10.1787/65cd27d4-en.

[20] Vendrell, M.; Johnston, S.K. Scaffolding critical thinking with generative AI: Design principles for integrating large language models in higher education. *Comput. Educ. Artif. Intell.* 2026, 10, 100572, DOI: 10.1016/j.caeai.2026.100572.



© 2026 by the author(s); licensee International Technology and Science Publications (ITS), this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)