

Discernment in generative AI assisted instructional design: An evaluation map and design heuristics for learning experience designers

Salil Sahadevan¹, Elizabeth Jacob², Manojkrishnan Champettil^{*3}, Deepak John Mathew⁴ and Sowmya Vivek⁵

¹Indian Institute of Technology Hyderabad, India & University Grants Commission, India

²New Delhi Institute of Management, New Delhi, India

^{3 5}EM Normandie Business School, Dubai, United Arab Emirates and EM Normandie MÉTIS Lab

⁴Department of Design, Indian Institute of Technology Hyderabad, India

Email: findsalil@gmail.com¹; lizjohn.tim@gmail.com²; djm@des.iith.ac.in⁴; svivek@em-normandie.fr⁵

*Correspondence: mchampettil@em-normandie.fr

Abstract

Existing research shows that generative artificial intelligence (GenAI) increases instructional design efficiency. However, there are concerns about pedagogical quality, accuracy and content bias. While prior studies address student passivity in evaluating AI content, the potential for designers to be passive in creating instructional materials remains largely unexplored. This article addresses design safeguards for AI-assisted course and learning-material development through two complementary tools: the Discernment Map, an evaluation checkpoint for instructional designers, and a set of design heuristics. The study is conceptual and design-oriented: both tools were drafted from the literature on AI fluency, automation bias and schema theory, and then refined through structured consultations with six experienced instructional designers in higher education. Though yet to be empirically validated, the practice-informed tools work at two levels: the Map operationalises human oversight of AI outputs across Product, Process and Performance dimensions for each AI-generated artefact, while the heuristics extend discernment into the design of the learning experience itself, so that learners must evaluate rather than accept AI outputs. Both are agile micro-governance tools that integrate critical judgment directly into design workflows. The approach is intended to protect long-term instructional quality against the collective erosion that unchecked AI adoption risks, and to support intentional engagement with AI that reduces automation bias.

Keywords: Generative AI, Instructional Design, Learning Experience Design, Discernment, Automation Bias, Design Heuristics, Human AI Collaboration, AI Fluency

JEL Classification: I21, I23, O33

Copyright: © 2026 by the authors. Licensee IJBM IEISS, New Zealand. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Generative artificial intelligence (GenAI) systems are common in course-authoring tools, learning-management platforms, and interactive interfaces. AI in education has long been expected to support teaching and learning (Luckin, 2018); GenAI tools now draft learning outcomes, content, assessments and feedback directly. While initial reports praise efficiency, these systems generate factually incorrect or disjointed material, sometimes misconstrued in context, yet often presented with persuasive fluency (Bender et al., 2021). When instructional designers uncritically ‘accept-and-paste’ such outputs, the immediate speed advantage comes at a later cost to teaching quality, epistemic reliability, and student well-being.

Students typically view AI use in learning favourably, especially with prior training or experience (Pellas, 2023). However, a positive outlook does not guarantee active engagement or critical use of AI material. Over-reliance, where students uncritically accept AI-generated recommendations or content, can harm critical thinking, decision-making, and analytical reasoning (Zhai et al., 2024). A recent work (Wang et al., 2023) suggests that simply following AI instructions, even with high participation, links to poorer performance and a weaker sense of community in AI-supported learning. A preprint EEG study of 54 participants reported that LLM users showed reduced neural connectivity and cognitive engagement compared with search-engine users and unassisted writers, weaker memory recall and a lower sense of ownership (Kosmyna et al., 2025).

Proposed remedies for passive consumption in AI-mediated learning include immediate feedback in AI tools (Cain, 2024; Pozdniakov et al., 2025), critical-thinking instruction (Cain, 2024), and training in effective AI use and prompt engineering (Cain, 2024). Less explored is the impact of decisions that instructional designers make during course creation, specifically the interplay between pedagogical assumptions and AI capabilities during learning material development.

Human expertise remains vital across studies for reviewing, refining, and placing AI-generated materials into context. AI-generated drafts can kickstart brainstorming but often lack the subtle, discipline-specific insights that seasoned educators provide (Dickey & Bejarano, 2024; Ullmann et al., 2024). AI-driven platforms can tailor content to individual learning preferences, thereby increasing engagement and perceived relevance. Still, the effectiveness of such personalisation depends on prompt quality, data accuracy, and ongoing human oversight (Dickey & Bejarano, 2024; Faghih Shojaei et al., 2025), the last of which is a preprint. AI-generated content is received more favourably in conventional courses than in unconventional ones (Kim & Koo, 2026), which demand deeper contextual and epistemic sensitivity.

Given the increasing reliance on AI-generated content in instructional settings, a structured set of design heuristics remains a gap. These heuristics would guide instructional designers in tackling two linked problems: their own cognitive biases during design, and the risk of passive content consumption by learners. If instructional designers use AI outputs without a critical eye, they risk embedding unexamined assumptions into learning materials. This could further entrench passive

consumption habits (Zhai et al., 2024). Design heuristics—cognitive strategies based on experience that aid decision-making in complex educational situations—offer a feasible way to address these issues (Bickle, 2024; Yilmaz et al., 2010; York & Ertmer, 2011). Such heuristics, derived from expert practice, have been shown to extend novice designers' exploration and judgment (Daly et al., 2012; Gray et al., 2019). Applied to AI-mediated design, they can direct designers to assess AI-generated materials critically and to promote learner independence.

Effective oversight hinges on discernment. This is the human ability to judge the accuracy of AI outputs, the validity of the reasoning of AI, and the appropriateness of its interactive behaviour (Dakan & Feller, 2025; Dakan et al., 2025). Oakley et al. (2026) argue from cognitive neuroscience that strong discernment relies on robust internal knowledge structures stored in long-term memory; without these structures, designers find it hard to spot subtle errors or shallow logic. Automation-bias research shows that users in that position become prone to 'automation bias', deferring to the system's output even when it is wrong (Goddard et al., 2012). Over-reliance on AI thus erodes the very capacity needed to evaluate it.

This mirrors, by analogy, the 'tragedy of the commons' described by Garrett Hardin (1968). In Hardin's account, rational herders, each seeking short-term gain, collectively ruin a shared pasture. In the learning-design domain, the 'commons' refers to the integrity of instructional materials and, by extension, public trust in educational materials and provisions. Every designer who cuts corners on evaluation for quick gains sees a private benefit. But when this multiplies across institutions, it leads to an influx of plausible yet shallow or biased content, diminishing the overall learning environment. The analogy has limits: instructional quality is not a finite stock that depletes with use; what accumulates is unexamined content. The relevant lesson is Ostrom's rather than Hardin's: commons can be sustained by rules that their users devise, adopt and monitor themselves, without external control (Ostrom, 1990). What is needed, therefore, is a framework light enough for rapid design cycles yet strong enough to maintain long-term instructional quality, and held by the designers themselves.

This study proposes practice-informed mechanisms to address the above problem: the Discernment Map and a set of design heuristics to guide AI-aware instructional practice. The Map transforms the abstract idea of discernment into concrete checkpoints and questions. Designers can use these after reviewing each AI-generated artefact. Using semi-structured interviews combined with observed walkthroughs, in which six experienced instructional designers worked on live AI-generated outputs, we analyse how designers currently assess AI outputs, situate where errors occur, and consider how internal knowledge structures shape judgment depth. The results, detailed in Section 4, present both the Discernment Map and the heuristics as practical micro-governance tools.

2. Literature Review

This section lays the groundwork for the Discernment Map and the proposed design heuristics within existing scholarship, defining key concepts and their connections. In this article, ‘instructional designer’ includes the learning experience designer (LXD) role. Claims that rest on a preprint, a practitioner document or a media source are identified as such in the text.

2.1 Generative AI in Instructional Design

Generative AI encompasses systems such as ChatGPT, Claude and Bard, all capable of producing human-like content (Farrelly & Baker, 2023). These tools have seen widespread use in learning design, automating tasks such as drafting outcomes, creating assessments, and generating media. While they offer efficiency and scalability, early studies also report issues with accuracy, dependability, and bias in generated outputs (Budhwar et al., 2023; Preiksaitis & Rose, 2023). For instructional design, frameworks such as ARCHED (Li et al., 2025) show how such tools can be embedded in a workflow, provided transparency and pedagogical rigour are upheld. A 2026 systematic review of 30 empirical studies identifies risk management, the educator’s evaluation of AI-generated content for accuracy, bias and ethical concerns, as one of five core functions of AI in instructional design, with educators as gatekeepers (Gagaza & Sarmiento, 2026); how that gatekeeping is to be done is not specified. The key takeaway remains that GenAI is increasingly integrated into instructional workflows, but its true worth and dependability rely on human oversight.

2.2 AI Fluency and the Role of Discernment

The AI Fluency Framework of Dakan and Feller (2025), taught in a course produced with Anthropic (Dakan et al., 2025), outlines four main competencies: Delegation, Description, Discernment, and Diligence. Among these, Discernment proves crucial for maintaining quality in AI-assisted design. It involves evaluating AI outputs, reasoning, and behaviour. This ability consists of three parts (Dakan et al., 2025): i) Product discernment assesses the correctness, coherence, and educational suitability of the output; ii) Process discernment judges the clarity and soundness of the underlying reasoning of AI; and iii) Performance discernment evaluates the interactive style of the system, its responsiveness, and its ability to adapt. The framework is a practitioner document rather than a peer-reviewed source, and these three areas have not yet been operationalised for everyday instructional workflows.

2.3 Automation Bias and the Risk of Overtrust

Automation bias refers to the tendency of users to overly trust algorithmic output, even when it is flawed; it is closely related to automation complacency, in which a trusted system is monitored less than its reliability warrants (Goddard et al., 2012; Parasuraman & Manzey, 2010). This problem worsens in GenAI due to its fluent, human-like output, which can hide errors (Bender et al., 2021). A journalistic account (Ovide, 2025) describes how professionals in law, journalism,

and navigation readily accept AI outputs in critical situations. This points to an urgent need for strong human-in-the-loop safeguards within educational settings.

2.4 Cognitive Schema Theory and Judgment

Oakley et al. (2026), drawing on schema theory, argue that internal knowledge structures guide how we perceive things and spot errors. When designers lack relevant schematic knowledge in a particular area, they are less able to detect AI-generated inaccuracies or flawed teaching approaches. A controlled experiment, reported in a preprint, found that complete reliance on GenAI for writing tasks reduced comprehension accuracy by about a quarter, and that participants with stronger prior knowledge benefited most from AI assistance (Ju, 2023). The need is to strengthen the internal knowledge structures of designers to address such cognitive vulnerabilities.

2.5 Operationalising Discernment in Instructional Workflows and the Need for Heuristics

A SWOT analysis of ChatGPT-generated course maps concluded that its outputs still require domain knowledge, fact-checking and human moderation for quality control (Choi et al., 2024). However, few studies offer a detailed, practical tool for evaluating AI outputs at each step of course creation. The CEAT framework of Peng et al. (2025) is a notable exception; it examines bias in AI-generated educational content using automated word-embedding analysis and reiterates the need for fairness in AI outputs. Section 5.2 compares the Discernment Map against these frameworks (Table 4).

Overall, the literature points to three strands when it comes to evaluating AI outputs in the context of instructional design: i) a clear need to evaluate not only AI content but also its reasoning and engagement quality; ii) the importance of strong existing knowledge when applying discernment; and iii) a scarcity of usable evaluation tools and concrete heuristic frameworks for instructional designers.

3. Methodology

This paper uses a conceptual, design-oriented approach, with expert consultation with instructional designers. The study aims to develop and refine a structured evaluation framework, the Discernment Map, and propose a set of design heuristics to guide critical assessment of generative AI outputs in instructional design contexts. Rather than testing hypotheses or producing broadly applicable empirical findings, our focus remains on conceptual clarity, functional design utility, and reasoning that aligns with real-world practice (Gregor & Hevner, 2013; Venable et al., 2016).

3.1 Research Objectives

This study has two objectives:

1. To develop the Discernment Map, an operational framework for evaluating AI outputs across Product, Process, and Performance dimensions.

2. To propose a set of design heuristics grounded in expert-informed learning experience design that foregrounds discernment in AI-mediated learning.

3.2 Research Design

The methodological process involved four stages: conceptual clarification, theoretical integration, framework drafting and heuristic formulation, and expert feedback from practitioners. Discernment was defined as a three-part construct based on the AI Fluency Framework (Dakan & Feller, 2025) and related theories. Relevant literature informed the positioning of discernment and the need for heuristic frameworks. Preliminary versions of the Discernment Map and the initial design heuristics were drafted by the authors from the three strands identified in Section 2, and designed to be light, flexible, and adaptable; the basis of each heuristic is recorded in the Reasoning column of Table 2.

For refinement, these drafts were shared with six experienced instructional designers from higher education: two lead learning designers (D1, D2), two professors active in curriculum design and working with SWAYAM, India's national platform for online courses (D3, D4), and two digital learning specialists (D5, D6). Each had at least five years of experience in instructional design and actively used generative AI tools. Each took part in one online session of 60 to 140 minutes combining a semi-structured interview, a walkthrough of the draft Map and heuristics, and an observed segment in which the participant worked on a live AI-generated output with the generative AI tools they used in their own practice, while one of the authors observed. No think-aloud protocol was imposed; participants commented as they worked, and their comments and the observed actions were recorded in the session notes.

The session notes were reviewed by the authors and grouped, first by the three discernment dimensions and then inductively within each, into themes describing challenges, current practices, and suggested changes; no coding software or inter-coder procedure was used. Each theme was then examined against the draft tools and, where it pointed to a gap, a revision was made. Table 3 in Section 4.3 traces each theme to the change it produced.

3.3 Expert Review

Validation in this study rests on theoretical coherence and practice alignment: each element of the framework and every heuristic links to prior research (Sections 2 and 4.2), and the consultations indicated that the Map and heuristics reflect real-world problems, decision paths, and the language used by instructional designers, and that both tools could be used with minimal orientation, whether standalone or within instructional design workflows.

3.4 Scope and Limitations

This study does not make broad empirical claims or test formal hypotheses. It is limited by its small sample and its focus on conceptual development. The interactions with designers served to refine the framework and heuristics, not to achieve representativeness. Consequently:

- Findings offer insights but are not exhaustive.
- The framework and heuristics remain proposed artefacts, awaiting future empirical validation.
- Their usefulness depends on context and assumes designers possess foundational knowledge in teaching methods and content.
- The process dimension of the Map relies on rationales produced by the AI system on request. Such rationales are model-generated accounts that may not correspond to the computation that produced the output (Turpin et al., 2023); the Map treats them as material for the designer's own verification, not as evidence of the system's reasoning.

4. Findings

The development and refinement of the Discernment Map and the design heuristics produced a set of findings related to AI-assisted instructional design. These stem from a synthesis of existing literature and structured feedback sessions with six experienced instructional designers who regularly use generative AI tools. The findings are arranged to correspond with the research objectives in Section 3.1, followed by the observations from the consultations that informed the refinement of both tools (Section 4.3).

The two tools operate at different levels and are complementary. The Discernment Map is designer-facing: it is applied to each AI-generated artefact during production and structures the designer's evaluation across the product, process and performance dimensions. The heuristics are learner-facing: they shape the design of the learning experience itself, so that learners meet AI outputs under conditions that require evaluation rather than acceptance. The Map protects the quality of what is produced; the heuristics protect the quality of what learners do with it.

4.1 Development of the Discernment Map

This study developed the Discernment Map, a structured, actionable framework that supports instructional designers in assessing AI-generated outputs across product, process and performance discernment (as defined in Section 2.2).

Table 1. The Discernment Map: evaluation focus, checkpoint questions and design actions for each AI-generated artefact, across the product, process and performance dimensions.

Dimension	Evaluation focus	Key questions	Design actions
Product	Quality and appropriateness of the AI-	▪ Is the information	✓ Correct factual or

discernment	generated output	accurate?	conceptual errors.
		▪ Is the content coherent and pedagogically relevant? ▪ Does it align with the learning goals?	✓ Revise for clarity, tone and alignment. ✓ Supplement with domain-specific inputs.
Process discernment	The rationale offered for the AI output: its logic, assumptions and structure	▪ Is the rationale logical and well supported? ▪ Are assumptions or steps missing? ▪ Can the sequence of design decisions be traced or explained?	✓ Request a concise rationale: the assumptions, evidence and design steps behind the output. ✓ Prompt for a step-wise, verifiable design rationale rather than for hidden internal reasoning. ✓ Reconstruct the reasoning independently, or scaffold it manually.
Performance discernment	Collaborative behaviour and adaptiveness of the AI tool	▪ Was the AI responsive to feedback? ▪ Was its tone appropriate for the learner group? ▪ Did it revise in meaning, or only rephrase?	✓ Re-prompt to test adaptiveness. ✓ Provide directive feedback. ✓ Switch models or interfaces if needed.

The Discernment Map works best as a repeated checkpoint during each design phase (e.g., after generating the content, assessment items, or lesson flow). It can be integrated into team workflows, faculty development sessions, or directly into AI-assisted authoring tool features. The framework

assumes designers possess existing subject knowledge; its goal is to support that knowledge with critical guidance, not to replace it.

A composite example from the consultations illustrates its use. A designer asks a generative AI tool to draft a learning outcome and a five-item quiz on climate change. At the product checkpoint, the outcome is well phrased but conceptually thin on greenhouse gas mechanisms, and the designer corrects it. At the process checkpoint, the designer asks for the assumptions and evidence behind the quiz and the sequence of design decisions; the rationale shows that two items presuppose a concept not yet introduced, so the designer re-sequences them. At the performance checkpoint, the designer asks for the tone to be revised for first-year students; the tool returns the same items with lighter wording, which the designer recognises as rephrasing rather than revision, and re-prompts with directive feedback before switching tools.

4.2 Proposal of Design Heuristics for AI-Aware Learning Experience Design

This study proposes a set of design heuristics grounded in expert-informed instructional design and learning experience design that foreground discernment in AI-mediated learning. These heuristics guide the design of learning experiences where AI is present but not dominant, ensuring that human judgment, intentionality, and epistemic depth are preserved.

Table 2. Design heuristics for AI-aware learning experience design: reasoning (with its source in the literature), guidance for the designer, and an example of use.

Heuristic	Reasoning	Guidance	Example
Design for Friction, Not Just Flow	AI tools often promote cognitive ease (auto-suggestions, summaries, completions), which can discourage deeper thinking (Oakley et al., 2026).	Avoid over-optimising for seamlessness; instead, build in deliberate pauses, prompts for reflection, or interpretive tasks that encourage learners to slow down.	After AI-generated content is presented, include a contrasting perspective or instruct learners to critique it using their own reasoning.
Surface the System-Logic	Learners need to understand how the AI arrives at its outputs to critically evaluate them; fluent output can conceal error and omission (Bender et al., 2021).	Make the logic of the AI, its limitations, and the provenance of its output visible within the learning experience.	Show how recommendations or summaries are generated, or provide metadata about source uncertainty or missing viewpoints.
Design for	The way learners	Treat prompt creation and	Include activities

Prompt Literacy	interact with AI (especially generative AI) depends heavily on their ability to create and interpret prompts and frame queries (Cain, 2024).	interpretation as a designable skill, not a hidden backend process.	where learners compare multiple prompt outcomes or reverse-engineer effective prompts.
Preserve Space for Ambiguity and Emergence	AI often resolves ambiguity prematurely or suppresses context-dependent details, whereas retention and understanding depend on the learner's own effort (Oakley et al., 2026).	Avoid over-curating content or rigidly pre-structuring outcomes; instead, allow learners room to encounter uncertainty and construct their own meaning.	In open-ended projects, resist auto-completion or scaffolding that diminishes creative freedom or intellectual struggle.
Design with Epistemic Pluralism	Most AI models tend to reproduce dominant paradigms and embedded bias (Peng et al., 2025); learners must encounter alternative ways of knowing.	Include knowledge sources, perspectives, and practices that AI may exclude or simplify.	Contrast AI-generated text with oral histories, lived experiences, or local and regional knowledge systems.
Calibrate Trust, Avoid Straight Adoption	Discernment relies on recognising when AI is helpful and when it is misleading or irrelevant; uncalibrated trust produces automation bias (Goddard et al., 2012; Parasuraman & Manzey, 2010).	Help learners and educators evaluate when to trust, question, or override AI outputs.	Use case-based learning to show examples where AI outputs led to error, bias, or unintended learning consequences.

4.3 Challenges and Practices in Evaluating AI-Generated Outputs

This subsection reports the observations from the consultations that informed the refinement of both tools. In each case the observation is stated first, followed by the literature with which it is consistent; Table 3 traces each observation to the revision it produced.

Surface plausibility in product discernment: In the consultations, participants noted instances where AI-generated content seemed convincing but was misleading. One participant, a lead learning designer, reviewed an AI-created course outcome about climate change; it was technically well phrased but lacked accurate conceptual detail on greenhouse gas mechanisms. This is consistent with the fluent but ungrounded output that Bender et al. (2021) describe and with earlier warnings about automation bias in other domains (Goddard et al., 2012); discernment prompts therefore need to be embedded during AI-content review, not applied as an after-the-fact correction.

Reasoning traceability in process discernment: Across all sessions, process discernment emerged as the most fragile aspect. Participants rarely asked the AI for the rationale behind its output or for its assumptions unless obvious errors appeared, and the reasoning behind instructional scaffolding or quiz generation was usually unclear even when the output seemed sound. Without a stated rationale, participants defaulted to superficial cues, raising the chance of logical inconsistencies in the teaching flow. In response, prompts for reasoning traceability were added to the Map, such as ‘State the assumptions and evidence behind this quiz and the sequence of design decisions.’ Such rationales are model-generated accounts that need not correspond to the computation that produced the output (Turpin et al., 2023); the Map treats them as material for the designer’s own verification, not as evidence of the system’s reasoning.

Iteration and responsiveness in performance discernment: Performance discernment, how the AI behaves during collaborative iteration, was considered important but often neglected. Participants noticed that some AI systems (e.g., Claude) responded more interactively than others, adjusting tone or depth in response to critique, but admitted that they rarely checked for this adaptability unless the tone was clearly mismatched. Several participants also reported ‘false responsiveness’, where outputs were merely rephrased rather than substantively changed: the interaction resembles collaboration without any change in meaning. This is consistent with a scoping review that notes the unwarranted confidence accompanying inaccurate generative AI outputs (Preiksaitis & Rose, 2023). To address this, the Map includes a responsiveness check (‘Did it revise in meaning, or only rephrase?’).

Across all three discernment dimensions, the participants’ own domain knowledge appeared to be the strongest influence on the quality of their discernment. Participants with more experience in the subject being designed, by role and by years of practice, identified flaws faster, suggested more precise re-prompts and showed greater confidence in their critiques. This matches accounts of expertise as well-organised knowledge structures (Chi, 2006) and calls to rebalance design training with retrieval practice, not just prompt engineering (Oakley et al., 2026).

Table 3. Feedback from the consultations and the revisions it produced in the Discernment Map and the design heuristics.

Observation or feedback from the consultations	Revision made
Participants asked for explicit guidance on how to challenge the assumptions built into AI outputs.	The heuristic ‘Surface the System-Logic’ was added to Table 2.
Participants rarely asked for the rationale behind quiz or scaffold outputs unless an error was obvious.	Rationale prompts were added to the process row of the Map, and the design action was worded to request assumptions, evidence and design steps rather than hidden reasoning.
Several participants reported ‘false responsiveness’: re-prompted outputs were rephrased but not substantively changed.	The check ‘Did it revise in meaning, or only rephrase?’ was added to the performance row of the Map.
Fluent but conceptually thin outputs, such as the climate-change learning outcome, passed initial review.	The Map was positioned as a checkpoint applied during the review of each artefact, rather than as an after-the-fact correction.

5. Discussion

The primary contribution of the study is twofold: the creation of the Discernment Map for instructional designers, and the proposal of a set of design heuristics for AI-aware learning experience design. Both tools, fully described in Section 4, offer actionable guidance at different levels: the Map to the designer evaluating each AI-generated artefact, the heuristics to the design of the learning experience that learners encounter.

These tools move beyond general principles or post-creation content reviews. They integrate directly into the instructional design process to support human judgment while working with AI in real time. Both are micro-governance tools in the sense introduced in Section 1: rules held and applied by the designers themselves, protecting long-term instructional quality without external rules or workflow disruption.

5.1 Discernment as Schema-Dependent Evaluation

The findings suggest that effective discernment is not merely about paying attention or knowing how to use a tool; it appears to depend on one’s existing knowledge structures (Chi, 2006).

Participants with well-organised internal knowledge spotted subtle inaccuracies, flawed reasoning, and inappropriate teaching tones more readily. This aligns with the argument of Oakley et al. (2026) that relying on GenAI bypasses the brain's prediction-error mechanisms, making it harder to 'feel' when something is wrong, and with automation-bias research such as Goddard et al. (2012). It is also consistent with a survey of 319 knowledge workers in which higher confidence in generative AI was associated with less critical thinking, and higher self-confidence with more (Lee et al., 2025).

In this context, the Discernment Map helps designers make up for the loss of unspoken expertise in environments where AI is heavily used. This points to a new necessity: any AI-fluency framework should include strategies to restore cognitive abilities, such as regular practice recalling information or 'critical thinking warm-ups' before using AI, to keep knowledge structures intact.

5.2 Addressing Overtrust and Surface Fluency

Consistent with Bender et al. (2021) and Preiksaitis and Rose (2023), this study observed that designers often placed too much trust in fluent AI outputs, even when these were logically flawed or pedagogically unsuitable. The Discernment Map counters this risk by embedding guided inspection routines at key points in the design cycle; the heuristics (Section 4.2), particularly 'Surface the System-Logic', reinforce it by prompting designers to consider the limitations of AI and to engage critically rather than accept blindly. A mixed-methods study of 70 instructional designers reports the same pattern from the practitioner side: generative AI is valued as a starting point for drafts, but its outputs require editing and verification before use (Luo et al., 2025).

The Map differs from a conventional quality-assurance checklist in what it inspects and when. A checklist inspects against fixed criteria such as alignment, clarity and accessibility. The Map inspects the judgment of the designer during production, and it covers the rationale offered for the output and the behaviour of the tool under iteration. Table 4 compares the Map with the existing frameworks. The Map and heuristics are a per-artefact checkpoint through which the designer exercises discernment. It is also a practical form of the gatekeeping role that Gagaza and Sarmiento (2026) identify but do not specify.

Table 4. The Discernment Map compared with existing frameworks for AI-assisted instructional design and AI fluency, by focus, stage of design, evaluator, and unit evaluated.

Framework	Focus	Stage of design	Evaluator	Major Evaluation Unit
ARCHED (Li et al., 2025)	A human-centred, transparent and collaborative workflow for AI-assisted	The whole design process.	The designer, guided by the framework's	The design workflow and its outputs.

	instructional design.		stages.	
GAIDE (Dickey & Bejarano, 2024)	Using generative AI to develop course content.	Content development.	The designer.	Course content.
CEAT (Peng et al., 2025)	Detecting bias in AI-generated educational content.	After generation.	Automated word-embedding analysis.	Mostly the textual output.
AI Fluency 4D framework (Dakan & Feller, 2025)	Competencies for working with AI, including discernment.	Any interaction with AI.	The individual user.	All outputs, process and behaviour, described at the level of a competency.
Discernment Map and the Heuristics (this study)	The designer's evaluation of each AI-generated artefact across product, process and performance.	During production, at each artefact.	The designer, per artefact.	All output, the rationale offered for it, and the tool's behaviour under iteration.

5.3 Practical Implications

This research has practical implications for training, edtech, and policy:

- **Instructional Design Training:** AI literacy programmes should integrate discernment practices, cognitive bias awareness, and judgment scaffolding strategies. They may also provide explicit instruction on how to apply the proposed design heuristics.
- **EdTech Platform Design:** Authoring environments should incorporate reflective checkpoints, such as toggles for 'Show rationale and assumptions' or 'Highlight assumptions', integrated with the logic of the Discernment Map. These platforms should build features that directly support the application of the design heuristics, such as tools to surface the AI's logic or encourage epistemic pluralism.
- **Institutional QA Policy:** Curriculum review bodies should consider discernment, as guided by the proposed heuristics, as a new dimension of instructional quality assurance in AI-rich design

workflows. This would complement existing metrics such as alignment, clarity, and accessibility.

6. Conclusion

Discernment in AI-supported learning design is neither optional nor instinctively clear. As GenAI becomes more integrated, human oversight must become more intelligent, not merely faster or more fluent. The Discernment Map and the set of design heuristics are steps towards that aim.

The AI-fluency literature treats discernment as a learnable competence (Dakan & Feller, 2025); in this study, its exercise depended on the expertise of the designer. Evaluating the rationale offered for an output, while often overlooked, is important for ensuring the logical accuracy of AI-generated instructional materials. Lightweight tools like the Discernment Map and the design heuristics proposed in this study can act as micro-governance tools, helping individuals take long-term pedagogical responsibility even amidst pressures for short-term efficiency.

While expert feedback indicated the immediate usefulness of the framework and its alignment with practice, future research should investigate: i) the inter-rater consistency of the Discernment Map and the application of the design heuristics; ii) their integration into actual design cycles; iii) their effect on learner outcomes and on how well users calibrate their trust in AI; and iv) how the framework and heuristics support collective discernment in team-based design settings or peer-review contexts.

References

- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445922>
- Bickle, J. T. (2024). *Exploration of pragmatic heuristic rendering within cognitive apprenticeship for remote teams: An exploratory Delphi study* [Doctoral dissertation, University of North Texas]. UNT Digital Library. <https://digital.library.unt.edu/ark:/67531/metadc2415984/>
- Budhwar, P., Chowdhury, S., Wood, G., Aguinis, H., Bamber, G. J., Beltran, J. R., Boselie, P., Cooke, F. L., Decker, S., DeNisi, A., Dey, P. K., Guest, D., Knoblich, A. J., Malik, A., Paauwe, J., Papagiannidis, S., Patel, C., Pereira, V., Ren, S., . . . Varma, A. (2023). Human resource management in the age of generative artificial intelligence: Perspectives and research directions on ChatGPT. *Human Resource Management Journal*, 33(3), 606–659. <https://doi.org/10.1111/1748-8583.12524>
- Cain, W. (2024). Prompting change: Exploring prompt engineering in large language model AI and its potential to transform education. *TechTrends*, 68(1), 47–57. <https://doi.org/10.1007/s11528-023-00896-0>

- Chi, M. T. H. (2006). Two approaches to the study of experts' characteristics. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 21–30). Cambridge University Press. <https://doi.org/10.1017/CBO9780511816796.002>
- Choi, G. W., Kim, S. H., Lee, D., & Moon, J. (2024). Utilizing generative AI for instructional design: Exploring strengths, weaknesses, opportunities, and threats. *TechTrends*, 68(4), 832–844. <https://doi.org/10.1007/s11528-024-00967-w>
- Dakan, R., & Feller, J. (2025). *AI Fluency Framework*. <https://aifluencyframework.org/>
- Dakan, R., Feller, J., & Anthropic. (2025). A closer look at discernment. In *AI fluency: Framework and foundations* [Online course]. Claude Academy. <https://academy.claude.com/courses/ai-fluency-framework-foundations/a-closer-look-at-discernment>
- Daly, S. R., Christian, J. L., Yilmaz, S., Seifert, C. M., & Gonzalez, R. (2012). Assessing design heuristics for idea generation in an introductory engineering course. *International Journal of Engineering Education*, 28(2), 463–473. https://www.ijee.ie/articles/Vol28-2/29_ijee2559ns.pdf
- Dickey, E., & Bejarano, A. (2024). GAIDE: A framework for using generative AI to assist in course content development. In *2024 IEEE Frontiers in Education Conference (FIE)* (pp. 1–9). IEEE. <https://doi.org/10.1109/FIE61694.2024.10893132>
- Faghih Shojaei, M., Gulati, R., Jasperson, B. A., Wang, S., Cimolato, S., Cao, D., Neiswanger, W., & Garikipati, K. (2025). *AI-University: An LLM-based platform for instructional alignment to scientific classrooms*. arXiv. <https://doi.org/10.48550/arXiv.2504.08846>
- Farrelly, T., & Baker, N. (2023). Generative artificial intelligence: Implications and considerations for higher education practice. *Education Sciences*, 13(11), Article 1109. <https://doi.org/10.3390/educsci13111109>
- Gagaza, M. A., & Sarmiento, C. P. (2026). Artificial intelligence in instructional design: A systematic review. *International Journal of Social Science and Humanity*, 16(1), 33–37. <https://doi.org/10.18178/ijssh.2026.16.1.1275>
- Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association*, 19(1), 121–127. <https://doi.org/10.1136/amiajnl-2011-000089>
- Gray, C. M., McKilligan, S., Daly, S. R., Seifert, C. M., & Gonzalez, R. (2019). Using creative exhaustion to foster idea generation. *International Journal of Technology and Design Education*, 29(1), 177–195. <https://doi.org/10.1007/s10798-017-9435-y>

- Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly*, 37(2), 337–355. <https://doi.org/10.25300/MISQ/2013/37.2.01>
- Hardin, G. (1968). The tragedy of the commons. *Science*, 162(3859), 1243–1248. <https://doi.org/10.1126/science.162.3859.1243>
- Ju, Q. (2023). *Experimental evidence on negative impact of generative AI on scientific learning outcomes*. arXiv. <https://doi.org/10.48550/arXiv.2311.05629>
- Kim, H., & Koo, T. K. B. (2026). The impact of generative AI on syllabus design and learning. *Journal of Marketing Education*, 48(1), 20–41. <https://doi.org/10.1177/02734753241299024>
- Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2025). *Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task*. arXiv. <https://doi.org/10.48550/arXiv.2506.08872>
- Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). 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* (pp. 1–22). Association for Computing Machinery. <https://doi.org/10.1145/3706598.3713778>
- Li, H., Fang, Y., Zhang, S., Lee, S. M., Wang, Y., Trexler, M., & Botelho, A. F. (2025). ARCHED: A human-centered framework for transparent, responsible, and collaborative AI assisted instructional design. *Proceedings of Machine Learning Research*, 273, 94–104. <https://proceedings.mlr.press/v273/li25a.html>
- Luckin, R. (2018). *Machine learning and human intelligence: The future of education for the 21st century*. UCL Institute of Education Press.
- Luo, T., Muljana, P. S., Ren, X., & Young, D. (2025). Exploring instructional designers' utilization and perspectives on generative AI tools: A mixed methods study. *Educational Technology Research and Development*, 73(2), 741–766. <https://doi.org/10.1007/s11423-024-10437-y>
- Oakley, B., Johnston, M., Chen, K.-Z., Jung, E., & Sejnowski, T. (2026). The memory paradox: Why our brains need knowledge in an age of AI. In M. Rangeley & N. Fairfax (Eds.), *The artificial intelligence revolution: Challenges and opportunities* (pp. 573–628). Springer. https://doi.org/10.1007/978-3-032-11814-1_22
- Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511807763>

- Ovide, S. (2025, June 3). You are hardwired to blindly trust AI. Here's how to fight it. *The Washington Post*. <https://www.washingtonpost.com/technology/2025/06/03/dont-trust-ai-automation-bias/>
- Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, 52(3), 381–410. <https://doi.org/10.1177/0018720810376055>
- Pellas, N. (2023). The influence of sociodemographic factors on students' attitudes toward AI-generated video content creation. *Smart Learning Environments*, 10(1), Article 57. <https://doi.org/10.1186/s40561-023-00276-4>
- Peng, J., Shen, W., Rao, J., & Lin, J. (2025). Automated bias assessment in AI-generated educational content using CEAT framework. In A. I. Cristea, E. Walker, Y. Lu, O. C. Santos, & S. Isotani (Eds.), *Artificial intelligence in education. Posters and late breaking results, workshops and tutorials, industry and innovation tracks, practitioners, doctoral consortium, blue sky, and WideAIED* (pp. 352–358). Springer. https://doi.org/10.1007/978-3-031-99264-3_44
- Pozdniakov, S., Brazil, J., Mohammadi, M., Dollinger, M., Sadiq, S., & Khosravi, H. (2025). AI-assisted co-creation: Bridging skill gaps in student-generated content. *Journal of Learning Analytics*, 12(1), 129–151. <https://doi.org/10.18608/jla.2025.8601>
- Preiksaitis, C., & Rose, C. (2023). Opportunities, challenges, and future directions of generative artificial intelligence in medical education: Scoping review. *JMIR Medical Education*, 9, Article e48785. <https://doi.org/10.2196/48785>
- Turpin, M., Michael, J., Perez, E., & Bowman, S. R. (2023). Language models don't always say what they think: Unfaithful explanations in chain-of-thought prompting. *Advances in Neural Information Processing Systems*, 36, 74952–74965. <https://doi.org/10.52202/075280-3275>
- Ullmann, T. D., Edwards, C., Bektik, D., Herodotou, C., & Whitelock, D. (2024). Towards generative AI for course content production: Expert reflections. *European Journal of Open, Distance and E-Learning*, 26(S1), 20–34. <https://doi.org/10.2478/eurodl-2024-0013>
- Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: A framework for evaluation in design science research. *European Journal of Information Systems*, 25(1), 77–89. <https://doi.org/10.1057/ejis.2014.36>
- Wang, X., Liu, Q., Pang, H., Tan, S. C., Lei, J., Wallace, M. P., & Li, L. (2023). What matters in AI-supported learning: A study of human-AI interactions in language learning using cluster analysis and epistemic network analysis. *Computers & Education*, 194, Article 104703. <https://doi.org/10.1016/j.compedu.2022.104703>



- Yilmaz, S., Seifert, C. M., & Gonzalez, R. (2010). Cognitive heuristics in design: Instructional strategies to increase creativity in idea generation. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 24(3), 335–355. <https://doi.org/10.1017/S0890060410000235>
- York, C. S., & Ertmer, P. A. (2011). Towards an understanding of instructional design heuristics: An exploratory Delphi study. *Educational Technology Research and Development*, 59(6), 841–863. <https://doi.org/10.1007/s11423-011-9209-2>
- Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments*, 11(1), Article 28. <https://doi.org/10.1186/s40561-024-00316-7>