computers

MDPI

Systematic Review

Generative AI in Higher Education Instructional Design: A Systematic Review of AI Roles, Educator Competences, and Professional Development Approaches

Natalia D. Spyropoulou * and Achilles D. Kameas

School of Science and Technology, Hellenic Open University, 26335 Patras, Greece; kameas@eap.gr
* Correspondence: nspirop@eap.gr

Abstract

Higher education (HE) educators have rapidly adopted generative artificial intelligence (GenAI) for instructional design (ID), yet evidence on this educator-facing use remains scattered. This PRISMA-guided qualitative systematic review synthesises 78 studies (2022–2026) to examine how AI-supported ID activity is distributed across ADDIE phases, functional AI roles, and practitioner versus pre-service contexts; the competences and human-oversight practices it requires of educators; and the professional development (PD) approaches documented for building them. Documented activity concentrates in the Design phase (95% of studies), while Implementation and Evaluation remain largely unexamined. Most studies (76%) apply no named competence framework; seven competence clusters emerge, with ethical awareness and bias detection recurrent yet weakly operationalised, suggesting a possible perceptual rather than declarative deficit. Human oversight is an explicit commitment in 58% of studies but examined empirically in only 21 (27%); the review consolidates five core oversight mechanisms plus two emerging ones, for which quantified evidence remains sparse and uneven. The corpus identifies six PD approach families, with embedded ethics checkpoints and performance-based assessment as the best-supported design features. The synthesis grounds an account of AI-supported ID in which the educator's design expertise, not the technology's generative fluency, remains the organising centre.

Keywords: generative artificial intelligence; instructional design; higher education; ADDIE; educator competences; Intelligent-TPACK; AI literacy; human oversight; professional development; systematic literature review

Check for updates

Academic Editor: Athanassios Jimoyiannis

Received: 30 July 2026
Revised: 18 October 2026
Accepted: 21 October 2026
Published: 22 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

The public release of large language model (LLM)-based generative artificial intelligence (GenAI) systems in late 2022 initiated a period of rapid, uneven, and largely unplanned integration of GenAI into higher education (HE) practice. Within months, educators across disciplines and national contexts used these tools to generate course content, draft learning objectives, design assessments, and create teaching materials [1,2]. The pace of adoption outstripped the production of evidence: institutions moved from blanket prohibition to managed-use policies without a reliable empirical basis for either decision [3], and instructional design (ID), the professional practice charged with ensuring pedagogical quality in technology-mediated learning environments, found itself navigating a landscape that had shifted faster than its knowledge base could accommodate.

The current moment is best understood as the third stage of a longer trajectory. Research on artificial intelligence in education (AIED) has a history spanning several decades, initially centred on intelligent tutoring systems, learner modelling, and adaptive platforms, technologies designed, built, and evaluated by specialists over long development cycles [4]. Reviews of this first wave consistently found the field dominated by student-facing applications: AI as tutor, recommender, feedback provider, or assessment engine [5,6]. The generative turn changed the terms of participation. Unlike earlier AIED systems, LLM-based tools require no specialist development, no institutional procurement, and no integration project: any educator with a browser can direct a general-purpose generative system toward any design task. This shift moved AI from the specialist's laboratory into the educator's everyday workflow and, in doing so, opened a genuinely new research object: the educator as user of AI in a professional design capacity.

Understanding this phenomenon requires a distinction that long predates AI: instructional design as distinct from teaching. Teaching is the enacted, interactive practice of instruction; instructional design is the systematic professional process that precedes and structures it: the analysis of learners and contexts, the specification of objectives, the selection of strategies, the development of materials, and the planning of assessment and evaluation [7]. Before a course meets its first student, its designer has taken dozens of consequential pedagogical decisions: what the learning outcomes will be and at what cognitive level; how content will be sequenced; which activities will serve which outcomes; how alignment between outcomes, activities, and assessment will be secured; what materials will be produced and in what media; and how evidence of effectiveness will be gathered. These decisions determine the quality of learning at scale, because they are inscribed in the course artefacts that every subsequent cohort encounters. When GenAI enters this process, it does not merely automate routine drafting and formatting; it participates in the decisions about objectives, sequencing, activities, and assessment that constitute instructional design itself. This is what makes the use of educator-facing AI analytically distinct from, and arguably more consequential than, the student-facing applications that have dominated the literature.

Yet the educator's perspective remains underrepresented. Zawacki-Richter et al. [5] first posed the question "where are the educators?" in 2019; reviews and meta-reviews as recent as 2024 indicate that it remains unresolved, continuing to find AIED research concentrated on learner-facing systems and learning analytics, with limited attention to how educators themselves use AI as a professional tool [6,8]. Prior syntheses fall into distinguishable families, reviews of AI in education broadly [4], of AI in higher education [5,6], of GenAI and ChatGPT specifically, the latter largely concerning its initial GPT-3.5-based release [9]; a related strand addresses AI literacy and educator competences [10,11]. These establish adoption patterns and general competence needs, but none takes educator-facing GenAI use in instructional design decision-making as its unit of analysis; that unit is precisely what the present review addresses, and it is why the existing syntheses, valuable as they are, cannot answer its questions. The imbalance is partly methodological: student outcomes are easier to instrument than design processes, and partly historical, since pre-generative AIED systems were typically specialist-built platforms rather than tools educators could apply directly to design work [4,5]. Neither reason survives the generative turn. Educators' use of GenAI in instructional design is now an empirical phenomenon of considerable scale, and while the evidence is growing, it has not been systematically mapped.

The stakes are higher in online and distance higher education. In campus-based teaching, design decisions can still be adjusted during instruction, through direct interaction between educator and learners; in distance education, the scope for such adjustment is far

more limited: courses reach large numbers of learners primarily through their designed materials, so the quality of the design decisions largely determines the quality of the learning experience [12]. Instructional designers in online and blended HE hold a distinct professional role precisely because of this institutional weight of design [13]. If GenAI is reshaping how design decisions are made, then evidence on this reshaping is most urgently needed in online and distance education.

This article addresses this evidence gap as the deliberately planned second step of a two-stage research programme. The first stage, our first work reported as a conference paper [14], applied PRISMA guidelines [15] to map the practice landscape in 20 practitioner studies, documenting where ADDIE phases include GenAI support, which design decisions it informs, and what functional roles the AI assumes relative to the educator. That mapping established that AI support is structurally concentrated in content production, and that, within those 20 studies, human oversight (the term employed by the EU AI Act and international guidance; Section 2.4) was widely endorsed but not empirically examined. These findings raise, but cannot answer, the questions that matter for institutions, educators, and regulators: what must the educator know and be able to do for this practice to be effective and responsibly applied, how is oversight actually exercised, and how can the necessary competences be built? This article addresses these questions based on an updated and expanded corpus from the same systematic review (Section 3). This study is a qualitative systematic review combining descriptive mapping with thematic synthesis: coded categories are quantified descriptively, and the qualitative material within them is synthesised thematically. The review can therefore support claims about what the reviewed literature documents, and how frequently, but not causal claims about the effectiveness of the practices described.

The article makes five contributions: an extension of the phase-and-role mapping across the updated corpus, with a systematic practitioner/pre-service comparison; a competence synthesis organised through the Intelligent-TPACK and AI literacy frameworks, revealing both the field's framework gap and its emergent competence structure; a typology of human oversight mechanisms that consolidates the five core mechanisms of the first study and extends them with two emerging mechanisms, with direct relevance to institutional governance and regulatory compliance; an evidence-based account of professional development approaches and their design features; and an integrative synthesis assembling the mapping, competence, oversight, and development findings into an integrated conceptual map of the field's current state, which also grounds subsequent theoretical work. The remainder of the article is organised as follows: Section 2 establishes the theoretical background and states the research questions; Section 3 details the methodology; Section 4 presents the results by research question; Section 5 discusses findings at empirical, theoretical, and policy levels; Section 6 presents the integrative synthesis; Section 7 outlines future research; and Section 8 concludes.

2. Background

2.1. Instructional Design in Higher Education

Instructional design is the systematic process through which learning goals are translated into instructional materials, activities, and assessments, and their effectiveness is evaluated [7]. Its disciplinary identity rests on the claim that effective instruction can be designed—that learning outcomes improve when decisions about objectives, strategies, media, and assessment are made deliberately, based on analysis and data rather than habit or improvisation. In HE, ID operates in two institutional configurations: as a professional specialisation (instructional designers supporting faculty, particularly in online and

blended provision [13]) and as a distributed academic practice (faculty designing their own courses, usually without formal ID training [1]).

Several models structure this process. The ADDIE model (analysis, design, development, implementation, evaluation) is the most widely referenced framework for systematic ID, organising the process into five iterative phases [7]. Backward Design begins from desired results and works backward through acceptable evidence to learning experiences [16]. Constructive alignment requires that intended outcomes, teaching activities, and assessment tasks form a coherent system [17]. Universal Design for Learning foregrounds accessibility, prescribing multiple means of engagement, representation, and expression from the outset [18]. The Learning Design tradition formalises designs as shareable, machine-readable artefacts [12], while Design Thinking imports empathy-driven, iterative prototyping methods into educational contexts [19,20].

ADDIE serves as this review's analytical mapping framework for three reasons. First, its five phases provide a complete, mutually recognisable map of the design process onto which heterogeneous studies can be coded regardless of the design model they themselves espouse—Backward Design and constructive alignment activities, for instance, code naturally into Design, and UDL commitments into Analysis and Development. Second, ADDIE is deliberately model-agnostic at the learning-theory level, avoiding the imposition of a specific pedagogical stance on the corpus. Third, and most consequentially for this review, ADDIE's iterative structure—the cycle from Evaluation back through Analysis into redesign—encodes the quality-improvement logic of systematic ID [7], making visible precisely which parts of that logic current AI support reaches and which it leaves unexamined.

2.2. Generative AI for Instructional Design

GenAI systems produce content (text, images, code) in response to natural language prompts, generating output by predicting successive tokens from probability distributions learned over large training corpora. Outputs are fluent and contextually plausible but not verified: they are susceptible to factual inaccuracies (hallucinations), to the amplification of biases embedded in training data, and to a systematic mismatch between surface plausibility and substantive accuracy [3,21]. For instructional design, these technical properties have a direct professional implication: they make systematic educator evaluation of AI output not an optional safeguard but a structural necessity.

Across the emerging literature, GenAI support has been documented for nearly every functional territory of ID work. In learner and context analysis, AI has been used to interpret diagnostic data for intervention planning [22]. In curriculum mapping and course structure, it has produced course maps, module sequences, and constructively aligned outlines [2,23,24]. In learning objective formulation, chatbots have generated, refined, and taxonomically aligned objectives with systematically evaluated quality [25,26]. In activity and assessment design, documented uses span lesson planning [22,27], field course design [28], mathematical problem posing [29], and rubric and question generation [25,30]. In material development, AI has produced culturally contextualised language materials [31], special education resources [32], simulation scenarios [33], and platform-ready course content including LMS code [2]. Applications in personalisation, feedback, and accessibility appear in adaptive course integration in blended contexts [13], AI-supported simulation debriefing [34], and accessibility-oriented resource adaptation [32]. The most recent literature extends this territory further: multi-agent systems draft complete course packages for faculty review [35], retrieval-augmented systems ground design support in institutional teaching knowledge [36], the reliability of LLMs in analysing lesson-plan quality is itself under systematic evaluation [37], and AI-generated feedback on design artefacts has been compared blind against expert feedback [38].

Two features of this evidence base motivate a systematic review. First, quality evaluations, where conducted, are sobering: only 38.3% of AI-generated learning objectives met all quality criteria without revision [25]; AI-generated objectives underperformed on specificity and achievability [26]; and plausible AI-designed field courses lacked the local ecological knowledge expert practitioners bring [28]. Second, the evidence is scattered across disciplines, methodologies, and terminologies, with no shared analytical structure, precisely the condition under which a phase-mapped, role-differentiated synthesis adds value.

2.3. Human–AI Collaboration and Oversight

Because this review asks how educators exercise oversight of AI-generated design content (RQ3), it requires a theoretical vocabulary for human—AI collaboration. Research on hybrid human–AI systems distinguishes configurations by the locus and timing of human control: in human-in-the-loop arrangements, a human validates or corrects each significant AI action before it takes effect; in human-on-the-loop arrangements, the AI operates autonomously while a human monitor and retains intervention authority [39]. Human-centred AI design argues that these are not fixed points but a design space: systems can be engineered for high automation and high human control simultaneously, with the goal of amplifying rather than replacing human capability [40]. Related concepts, such as human agency (the capacity to direct the process and own its outcomes), AI augmentation (AI extending human performance on a task the human still owns), and AI automation (AI performing the task in the human's stead), provide the analytical resolution needed to describe what actually happens when an educator designs with GenAI.

In instructional design, these concepts frame the educator as the accountable decision-maker in a hybrid configuration: the designer specifies intent, the AI generates candidate design content, and the outcome's quality depends on the oversight mechanisms through which the designer evaluates, corrects, and adopts or rejects that content. Hybrid-intelligence research in education argues that such configurations succeed when control transitions between human and machine are deliberately designed rather than left implicit [39]. Whether the emerging practice of AI-supported ID exhibits such deliberate design—or whether oversight remains an unexamined assumption—is an empirical question this review addresses directly. Recent literature has begun to address this empirically, from configurable human-involvement modes in multi-agent design systems [35] to evidence that critical evaluation of AI output shapes teacher agency [41].

2.4. Responsible AI in Education: The Policy Context

The oversight question is not merely academic; it is the central demand of the international policy architecture now forming around AI in education. UNESCO's guidance on generative AI in education calls for a human-centred approach in which AI augments rather than replaces educator judgement, and for institutional validation of AI outputs used in teaching [3]. The OECD's digital education analyses similarly position human oversight and effective governance as preconditions for trustworthy educational AI [42]. The European Commission's ethical guidelines for educators articulate requirements for human agency and oversight, transparency, fairness, and accountability in AI use for teaching and learning [43]. Most consequentially, the EU Artificial Intelligence Act establishes a binding, risk-based regulatory regime in which several educational uses of AI are classified as high-risk, triggering explicit obligations of human oversight and transparency [44]. Across these instruments, a consistent expectation emerges: humans who deploy AI in educational processes must be competent to supervise it. For the specific professional practice of instructional design, that competence is exactly the territory of RQ2.

2.5. Educator Competence Frameworks

Several frameworks describe the competences educators need for technology-mediated and AI-mediated teaching. The TPACK framework conceptualises effective technology integration as the intersection of technological, pedagogical, and content knowledge [45]; it remains the field's most widely applied lens but predates AI whose outputs are probabilistic, biassed, and agentive. DigCompEdu, the European framework for educators' digital competence, organises 22 competences across six areas, from professional engagement to facilitating learners' digital competence, and provides the structural template for European competence policy [46]. The AI literacy framework specifies a developmental progression—know and understand, use and apply, evaluate and create, plus ethical awareness—applicable to educators and learners alike [10]. Intelligent-TPACK extends TPACK with AI-specific knowledge domains and, critically, embeds a four-facet ethics strand (transparency, fairness, accountability, inclusiveness) within each domain, treating ethical competence as distributed and iterative rather than appended [11]. At the policy level, UNESCO's AI Competency Framework for Teachers defines five competence dimensions across progression levels, giving institutions a reference for professional development planning [47]. Most recently, project-level frameworks have begun translating these general schemes into role-specific operational guidance for HE stakeholders, including the ADMIT framework for the ethical use of GenAI in higher education [48].

Collectively, these frameworks describe what educators should know about AI in general. However, none was constructed for the specific professional practice this review examines: using GenAI in instructional design decision-making. Whether and how the empirical literature draws on these frameworks, and what competence structure emerges when it does not, is the substance of RQ2.

2.6. Research Questions

Against this background, four research questions guide the synthesis. They are ordered hierarchically: RQ1 establishes the descriptive foundation on which the remaining questions build; RQ2 and RQ3 examine what the mapped activity requires of educators; and RQ4 addresses how those requirements can be met.

RQ1: How is documented, AI-supported instructional design activity distributed across ADDIE phases and functional AI roles (as defined in Section 3.3), and how does this distribution differ between practitioner and pre-service teacher education contexts within HE? RQ1 deliberately pairs a mapping component with a comparative component, because the comparison is interpretable only against the mapping; the two are reported in sequence in Section 4.1.

RQ2: What competence frameworks are applied in the reviewed literature, and what competence clusters emerge from the documented practices? Frameworks (researcher-applied theoretical scaffolds) and clusters (empirically derived capability groupings) are conceptually distinct objects; they are therefore analysed separately, in Sections 4.2.1 and 4.2.2, respectively, and are not compared with each other.

RQ3: How is human oversight of AI-generated instructional content operationalised in educator practice, as documented in the reviewed literature?

RQ4: What professional development approaches for building these competences are documented, and what design features characterise the approaches with the strongest outcome evidence?

3. Materials and Methods

3.1. Search Strategy

The review followed the PRISMA 2020 statement [15]. Accordingly, the review specifies explicit eligibility criteria (Section 3.2), documents all information sources and the full search strategy (Section 3.1), applies a two-stage screening process with tagged exclusion reasons (Section 3.2), uses structured data extraction and coding (Section 3.3), and reports the complete study flow in Figure 1. The review was not registered and no separate protocol was prepared. Electronic searches covered three databases selected for complementary coverage: ERIC, as the reference database for education research, ensuring recall of pedagogically framed studies; and Scopus and Web of Science, as the two major curated multidisciplinary indexes, ensuring coverage of education-relevant work published in computing, health sciences, and engineering venues, where a substantial share of GenAI-in-education research appears. The search covered the period from November 2022, marking the public release of LLM-based GenAI systems, to October 2026; the search was first run in February 2025 and re-run with identical strings in all three databases on 16 October 2026, restricted to records published from 2025 onwards in order to target the literature added since the first execution. The three databases were selected because they jointly cover the education-research core (ERIC) and the curated multidisciplinary indexes in which GenAI-in-education work concentrated during the review window (Scopus, Web of Science); work indexed only in ACM Digital Library, IEEE Xplore, or discipline-specific databases may nonetheless be missed, and this is acknowledged as a limitation (Section 5.4). The full database-specific search strategies, fields, filters, and search dates are provided in the Supplementary Materials (Table S1).

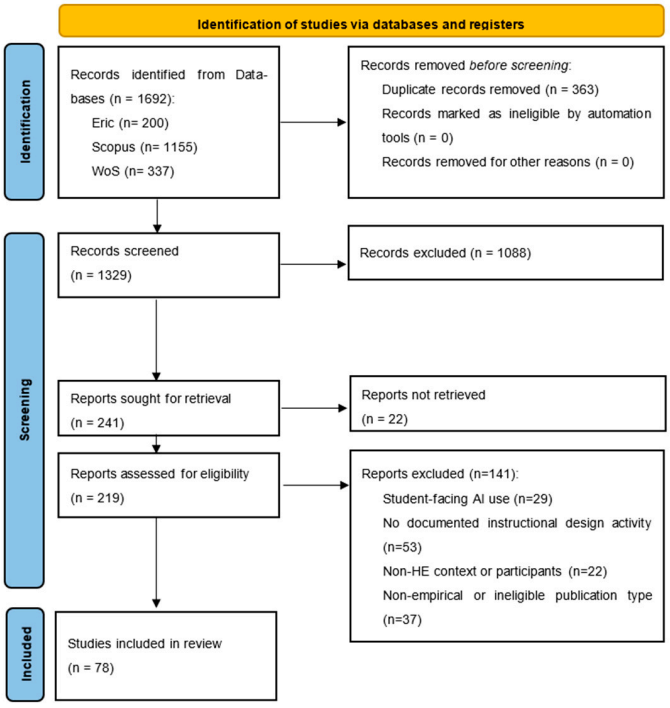


Figure 1. PRISMA 2020 flow diagram of the study selection process (search period November 2022–October 2026; 78 included studies).

The search string combined three concept blocks with Boolean operators: ("generative AI" OR "large language model*" OR "ChatGPT" OR "GPT") AND ("instructional design" OR "course design" OR "lesson planning" OR "curriculum design" OR "learning design") AND ("higher education" OR "university" OR "faculty" OR "educator*").

3.2. Eligibility Criteria and Selection

Records were screened against six inclusion criteria: (1) HE educators or pre-service teachers enrolled in HE institutions as participants; (2) AI used as an educator-facing ID tool; (3) educator ID activity as the studied phenomenon; (4) an HE institutional context; (5) publication in English; and (6) publication as a peer-reviewed journal article or conference paper (books and book chapters were excluded). The pivotal construct, educator-facing, means the educator uses the AI as a professional design instrument. Thus, a study in which an instructor uses a chatbot to draft learning objectives, generate a course map, or produce simulation scenarios is educator-facing and eligible; a study in which students use a chatbot to support their own learning, or in which an AI system delivers instruction directly to learners, is student-facing and ineligible, even if teachers appear as informants. The related distinction between "teaching with AI" (AI as the educator's design tool) and "teaching about AI" (AI as subject matter) was applied consistently: courses that merely taught AI content, without the educator using AI in the design process, did not meet criterion (2). Exclusion criteria comprised technology-acceptance and perception studies without design activity, student-outcome studies, opinion pieces without documented practice, and administrative AI applications.

We pooled the records from both search phases (Section 3.1; per-phase counts in Supplementary Table S1) and screened them against the same eligibility criteria. We exported the 1692 retrieved records, identified and removed duplicates (n = 363), and retained 1329 unique records. These were uploaded to Rayyan to organise the screening process; the platform was used solely for categorisation, tagging, and the documentation of screening decisions, and no automated or AI-assisted classification functions were employed. Exclusion tags recorded the reason for each removal, preserving an auditable record of decisions at each stage. Title-and-abstract screening excluded 1088 records falling outside the review's scope, principally literature and systematic reviews, studies of non-HE cohorts, perception surveys without documented design activity, studies centred on students rather than educators, reports of AI system development, and ineligible publication types. We sought the full texts of the remaining 241 reports; we could not retrieve 22, and we assessed 219 for eligibility. Of these, 141 were excluded: student-facing rather than educator-facing AI use, no documented instructional design activity, non-HE contexts or participants, or non-empirical or ineligible publication types, yielding 78 included studies (Figure 1). This steep reduction is expected for a review whose phenomenon is deliberately narrow within a rapidly expanding field: the search string necessarily also captures the much larger student-facing literature, which we explicitly exclude, along with reviews and perception studies without documented design activity. The first author performed screening at both stages. The first author flagged borderline and uncertain records (n = 22) and resolved them through discussion with the second author until consensus was reached, using the participant's institutional role and the study's declared phenomenon as deciding criteria; the second author also reviewed the inclusion decisions of the second search phase. Screening thus rested on a single reviewer, supported by these consensus and review steps; the implications are considered in Section 5.4.

The 78 included studies form two analytic sets. Forty-six (46) position practising HE educators, faculty members, instructional designers, and researcher-educators as the studied practitioners (RQ1–RQ3); 20 of these constituted the mapping corpus of the first study [14] and are identified in Section 4.1. The remaining 32 are situated in pre-service teacher education, where the studied phenomenon is competence development through design activity rather than professional design practice (RQ1–RQ4). Their inclusion follows directly from the participant criterion: what qualifies a study is that its participants are situated in higher education and perform instructional design activity there, not the profes-

sional status of the designer. Their evidentiary value, however, concerns what designing with AI develops rather than how professionals practise. Accordingly, all 78 studies inform the mapping, competence, and oversight analyses (RQ1–RQ3), while the pre-service set carries the principal professional development evidence (RQ4). We report the two sets jointly and separately: Section 4.1 presents the full-corpus mapping with its breakdown by set, and we treat the comparison between them as a finding in its own right (Section 4.1).

3.3. Data Extraction and Coding

Data were extracted into a structured spreadsheet organised in four blocks: bibliographic information; design context; AI role and human oversight; and study characteristics. We extracted the AI tool used and, where reported, its version for each study; complete study-level coding is reported in the Supplementary Materials (Table S3). Coding was performed by the first author and reviewed in periodic discussion sessions with the second author; ambiguous categorisations and coding disagreements were resolved by consensus. Coded entries were cross-checked against the original texts of the studies before being finalised, and the completed extraction file was reviewed in full by the second author; this collaborative review process was intended to limit the risk of individual coding bias and to strengthen the reliability of the synthesis. The MMAT appraisal informed the confidence placed in the synthesised findings rather than study inclusion or exclusion. Analysis combined descriptive quantification of the coded categories, frequencies and cross-tabulations, reported in Section 4, with thematic analysis of the extracted qualitative material, proceeding through familiarisation with the extracted data and full texts, initial coding, the grouping of codes into candidate themes, and iterative review and refinement of themes until the final categories reported in Section 4 were defined. Methodological quality of the included studies was appraised with the Mixed Methods Appraisal Tool (MMAT, 2018 version), applying the design-appropriate criteria set to each study; per-study appraisals are provided in Supplementary Table S2 and a confidence assessment of the principal findings in Supplementary Table S4, whose ratings are the authors' judgement across methodological quality (MMAT), number of supporting studies, consistency across studies and analytic sets, and directness to the review question, with the rating categories defined in the table note. Coding reliability rests on the consensus procedure described above rather than on independent double-coding, and no inter-rater agreement statistic is therefore reported (Section 5.4); to allow independent verification, every coding decision is reported per study in Supplementary Table S3. Throughout the article, "reported" denotes author self-description, "measured" denotes quantified assessment, "evaluated" denotes structured qualitative assessment, and "inferred" denotes categories derived analytically from documented practice. Coding proceeded as follows.

ADDIE phase coding. Each study was coded for the phase(s) in which AI-supported design activity was empirically documented, permitting multiple phases per study. Coding required documented activity, not incidental mention: a study was coded for Analysis only where AI demonstrably supported learner, context, or needs analysis (e.g., interpreting diagnostic reading data for intervention planning [22]), not where the authors merely discussed the phase. Studies were coded against ADDIE regardless of the design model they espoused; Backward Design and constructive-alignment activities were mapped to their functionally equivalent phases.

AI role coding. The three AI roles emerged from the analysis itself: an initial open-coding pass over the AI-use descriptions of the first ten studies produced candidate functional categories, which stabilised through iterative cross-study comparison into three: Generative (AI produces design content at the educator's direction), Scaffolding (AI structures the educator's own design reasoning), and Evaluative (AI assesses the quality of

design artefacts). The three-role scheme, first reported in the companion paper [14], was then applied deductively to the full corpus, with roles coded non-exclusively. For example, a study reporting that educators used ChatGPT to produce a complete course structure and iteratively critique their prompts was coded as Generative and Scaffolding.

Oversight and competence coding. We coded human oversight at two levels: an ordinal level (explicit commitment, moderate, minimal/none) and a mechanism level, extracting verbatim descriptions of how oversight was exercised, which we then clustered into the typology reported in Section 4.3. Each study was further coded for whether oversight was examined empirically, defined as the systematic collection and analysis of data on human oversight of AI outputs: criterion-based rating of AI-generated outputs by educators or experts, comparison of oversight or human-involvement configurations, or analysis of logged review, revision, or rejection of AI outputs (Supplementary Table S3). Competence coding (RQ2) used the Intelligent-TPACK domains [11] and AI literacy progression [10] as sensitising lenses while allowing clusters to emerge thematically; each study was additionally coded for whether it applied any named competence framework. For this count, we distinguished named competence and teacher-knowledge frameworks (TPACK and its AI-specific extensions, AI literacy frameworks, and models of pedagogical content knowledge) from task-specific prompting methods and broader learning theories, which we reported separately. PD coding (RQ4) captured intervention duration, structure, assessment method, and documented outcomes. We coded iteratively, flagging ambiguous codes and resolving them through full-text re-reading. Table 1 presents the characteristics of the included studies.

Table 1. Characteristics of the 78 included studies (2022–2026), ordered alphabetically.

Study	Focus	Country	Participants	Design	Set ††
Aguş & Mutlu (2026) [49]	Multimodal ELT materials, 2 DBR cycles	Turkey (Anadolu)	Instructors/researchers; 639 + 577 student outcomes; 4–5 instructor interviews	Mixed methods	Practitioner (RQ1–3)
Akpinar & Kutbay (2026) [50]	Prompting, SRL and AI literacy in lesson planning	Turkey	Pre-service teachers * (n = 63)	Quasi-experimental + prompt coding; mixed	Pre-service (RQ1–4)
Aldemir et al. (2025a) [51]	Emergent AI-TPACK competences in a two-week AI literacy module	USA	Pre-service teachers * (n = 29)	Exploratory case study; mixed methods	Pre-service (RQ1–4)
Aldemir et al. (2025b) [52]	Intelligent-TPACK in a three-week teacher preparation module	USA/Turkey	Pre-service teachers * (n = 25)	Convergent mixed methods	Pre-service (RQ1–4)
Aydin Gunbatar et al. (2025) [53]	Lesson planning (acids/bases, CoRe format)	Turkey (Van)	6 of mixed experience: 2 senior PSTs + 1 induction + 2 teachers + 1 teacher educator	Qualitative	Practitioner (RQ1–3)
Cai et al. (2025) [54]	Lesson plan critiques → revision	China/Singapore	48 PSTs (24 + 24 control), Jiangnan + NIE	Quantitative randomised controlled (+ENA)	Pre-service (RQ1–4)
Chen & Osman (2026) [55]	DeepSeek-supported ID and TPACK/CP development	China	Pre-service physics teachers * (n = 130)	Quasi-experimental	Pre-service (RQ1–4)
Chick & Morello (2026) [56]	Claude-based agentic rapid ID; UDL alignment	USA	Faculty designer (n = 1; doctoral programme, HSI)	Single-case; blinded rubric rating	Practitioner (RQ1–3)
Choi (2025) [57]	ChatGPT in 5E Earth-science lesson design	South Korea	Pre-service Earth-science teachers * (n = 8)	Qualitative case study	Pre-service (RQ1–4)
Choi et al. (2024) [23]	Generative AI for instructional design: SWOT † analysis of course map creation	USA	Instructional-design researchers (n = 4)	Practitioner inquiry (qualitative)	Practitioner (RQ1–3)

Table 1. *Cont.*

Study	Focus	Country	Participants	Design	Set ††
Conklin et al. (2024) [2]	Full online course creation with ChatGPT (Backward Design)	USA	Instructional-design faculty (n = 3)	Case study	Practitioner (RQ1–3)
Crowther et al. (2025) [25]	Chatbot-generated learning objectives: quality evaluation	USA	University instructors (4 courses)	Comparative evaluation	Practitioner (RQ1–3)
da Silva (2025) [58]	Course on AI for L&D practitioners—4 iterations	USA	Faculty design team (Boise State)	Design case (IJDL); Learning Engineering	Practitioner (RQ1–3)
Daniels et al. (2025) [33]	AI-enhanced game-based learning for project leadership	Ireland	Faculty designers; students	Design-led qualitative	Practitioner (RQ1–3)
Davis (2025) [59]	ChatGPT-supported flipped course design	—	HE faculty (physiology)	Design case (IJDL)	Practitioner (RQ1–3)
Dilek et al. (2025) [60]	AI literacy through critical co-discovery	USA	Educators in AI-in-education course	Illustrative case study	Pre-service (RQ1–4)
Doyle et al. (2025) [26]	AI-generated versus human-crafted learning objectives	USA	Computing instructors (evaluators)	Comparative evaluation	Practitioner (RQ1–3)
Faith et al. (2025) [61]	PBNJ critical-friend workflow for DI planning	Canada	Faculty (n = 2) + pre-service teachers * (n = 5); survey n = 147	Collaborative self-study; mixed	Pre-service (RQ1–4)
Farrag et al. (2026) [62]	Four-LLM pharmacotherapy simulation generation	—	Expert evaluators (artefact eval)	Comparative evaluation	Practitioner (RQ1–3)
Fernández-Sánchez et al. (2025) [63]	27 assessment rubrics	Spain	120 pre-service (Primary Ed, A Coruña) + instructors	Qualitative (action research)	Pre-service (RQ1–4)
Fu (2026) [64]	PE lesson plans with physiological load	China	60 pre-service PE teachers (randomised, with control group)	Quantitative randomised controlled	Pre-service (RQ1–4)
Ghouali (2025) [65]	Teacher evaluation of ChatGPT lesson plan (EFL)	Algeria	University EFL writing teachers (n = 15)	Artefact evaluation; questionnaire	Practitioner (RQ1–3)
Gold et al. (2026) [66]	Prompt histories in curriculum design tasks	USA	Pre-service teachers * (n ≈ 70)	Qualitative artefact/log analysis	Pre-service (RQ1–4)
Gültekin et al. (2026) [67]	CLO writing/evaluation (Bloom-aligned)	Turkey (Kocaeli)	22 faculty (authentic use) + 3 expert raters	Mixed methods	Practitioner (RQ1–3)
Guo & Chen (2026) [68]	GenAI in FL teaching preparation	—	University FL instructors (n = 13)	Qualitative interviews	Practitioner (RQ1–3)
Hauk & Soujon (2026) [37]	LLM reliability in evaluating written lesson plans	Germany	32 PST lesson plans; LLM as evaluator	Explanatory sequential mixed methods	Practitioner (RQ1–3)
Husain (2024) [69]	ChatGPT in computer programming instruction	Jordan	IT faculty (n = 12)	Qualitative interviews	Practitioner (RQ1–3)
Jang et al. (2025) [34]	Question-centred prompt engineering for nursing simulation debriefing	South Korea	Nursing instructor; students (n = 28)	Model development; pre–post mixed methods	Practitioner (RQ1–3)
Jiang et al. (2026) [38]	Instructional designs in an ID course; feedback from 3 sources	China (CCNU)	110 pre-service physics teachers	Mixed methods	Pre-service (RQ1–4)
Jou et al. (2025) [70]	Generative AI in teacher education: TPACK ‡ growth	Taiwan	Pre-service teachers * (vocational)	Qualitative thematic analysis	Pre-service (RQ1–4)
Kaczorowski et al. (2024) [32]	AI integration in special education teacher preparation	USA	Teacher-preparation faculty (n = 4)	Practitioner report	Practitioner (RQ1–3)
Kerneža (2023) [71]	Cognitive skills for chatbot use in education	Slovenia	Pre-service teachers * (n = 58)	Mixed methods (self- and external assessment)	Pre-service (RQ1–4)

Table 1. *Cont.*

Study	Focus	Country	Participants	Design	Set ††
Kerr & Kim (2025) [27]	Pre-service teachers' GenAI use for lesson planning	South Korea	Pre-service teachers * (English)	Qualitative case study	Pre-service (RQ1–4)
Kessner & Kaka (2025) [72]	Social studies lesson plans—quality evaluation	USA	Teacher educators as evaluators	Quantitative descriptive	Practitioner (RQ1–3)
King & Yan (2025) [73]	Designing a graduate GenAI-integrated course	USA	Faculty + doctoral designer	Design case (IJDL)	Practitioner (RQ1–3)
Kohnke et al. (2025) [74]	Microlearning and generative AI in teacher education	Hong Kong	Pre-service teachers * (n = 19)	Qualitative case study	Pre-service (RQ1–4)
Kumar et al. (2024a) [13]	Instructional designers and GenAI in online/blended HE	USA	Instructional designers (n = 15)	Qualitative interviews	Practitioner (RQ1–3)
Kumar et al. (2024b) [24]	ChatGPT for natural sciences course design: SWOT †	USA	Natural-sciences instructors (n = 5)	Qualitative workshop reflections	Practitioner (RQ1–3)
Leinonen et al. (2026) [75]	AI-generated slides across five GenAI tools	Finland/Canada	Computing educators (n = 3)	Case study; artefact evaluation	Practitioner (RQ1–3)
Li, H. (2026) [76]	DeepSeek-supported inquiry in multiplayer VR for ID skills	China	Pre-service teachers * (n = 47: 23 + 24 control)	Quasi-experimental with control	Pre-service (RQ1–4)
Li, T. et al. (2025) [77]	Inclusive STEM teaching preparation	China (SCNU)	Student teachers (2 studies, 2 years)	Mixed methods	Pre-service (RQ1–4)
Lin et al. (2025) [78]	Reading materials development (one semester)	China	2 university L2 Chinese teachers	Qualitative	Practitioner (RQ1–3)
Liu et al. (2026) [79]	ChatGPT courseware generation with concept maps	China	Pre-service teachers * (n = 91: 46 + 45 control)	Quasi-experimental with control; mixed	Pre-service (RQ1–4)
Lorenz & Kikkas (2025) [80]	ChatGPT in modular IT-ethics course redesign	Estonia	University instructors (n = 2)	Longitudinal practice study	Practitioner (RQ1–3)
McDonald et al. (2025) [81]	GAI + voice cloning in online course development	USA	Course development team (faculty + IDs)	Mixed methods (grades, evals, interviews)	Practitioner (RQ1–3)
Moorhouse et al. (2025) [82]	Refining lesson plans with task-specific prompting	Hong Kong	25 PSTs, MEd ELT (HKBU)	Qualitative	Pre-service (RQ1–4)
Moundridou & Matzakos (2025) [83]	WebQuest design (LBD framework)	Greece (ASPETE)	Pre-service mechanical engineering educators	Mixed methods	Pre-service (RQ1–4)
Ndlovu et al. (2026) [84]	AI-supported lesson planning and adaptive reasoning	South Africa	Pre-service science teachers * (n = 19)	Within-participant comparative; mixed	Pre-service (RQ1–4)
Oh (2025) [85]	MATH41 and generative AI in mathematics teacher education	South Korea	Pre-service teachers * (n = 24)	Mixed methods (journals; survey)	Pre-service (RQ1–4)
Okulu & Muslu (2024) [86]	Course plan development with ChatGPT	Turkey	Researcher-educators (n = 2)	Document analysis of AI dialogue	Practitioner (RQ1–3)
Panke (2025) [87]	Research + teaching + ID routines (5 months)	USA (UNC)	1 instructional designer/educator	Qualitative (autoethnography)	Practitioner (RQ1–3)
Park et al. (2026) [88]	Microlearning faculty PD with Copilot (institutional Enterprise Data Protection licence, summer 2024; model version not reported)	USA	HE faculty (Penn State PD)	PD intervention; mixed	Practitioner (RQ1–3)
Rodríguez Jerez et al. (2025) [30]	GenAI and 3D immersive environments in competency-based HE	Colombia	Aeronautical-school instructors	Mixed methods	Practitioner (RQ1–3)
Santiago (2026) [89]	Teaching strategies for environmental science	Philippines	College instructor (CAR) + students as data	Qualitative (action research)	Practitioner (RQ1–3)

Table 1. *Cont.*

Study	Focus	Country	Participants	Design	Set ††
Schilhabel (2025) [90]	Redesign publisher slide decks	USA (UW Oshkosh)	1 faculty (IS capstone, 7 sections)	Qualitative (DSR case)	Practitioner (RQ1–3)
Segal & Biton (2024) [29]	GenAI problem posing and TPACK ‡ development	Israel	Pre-service teachers * (maths; n = 15)	Mixed methods questionnaire	Pre-service (RQ1–4)
Seufert et al. (2025) [91]	Intelligent-TPACK through AI assistance	Switzerland	Pre-service teachers * (business; n = 12)	Multi-method	Pre-service (RQ1–4)
Sha et al. (2026) [41]	Critical evaluation of GenAI outputs and teacher agency	Australia	University educators (n = 13, computing)	Mixed-methods trace analysis	Practitioner (RQ1–3)
Squalli Houssaini et al. (2024) [20]	Design thinking and GenAI for medical curriculum	Morocco	Medical faculty	Case study	Practitioner (RQ1–3)
Sun & Huang (2026) [92]	Collaborative STEM lesson plan refinement	China	12 PSTs (6 dyads), ECNU	Mixed methods	Pre-service (RQ1–4)
Suresh et al. (2025) [93]	GenAI in learning design practice—online course	USA	University design team (Iowa State)	Design case (IJDL)	Practitioner (RQ1–3)
Tran et al. (2025) [1]	Practitioner perspectives on ChatGPT and course design	Multi-country	Lecturers; educational specialists (n = 22)	Qualitative interviews	Practitioner (RQ1–3)
Tseng & Lin (2024) [94]	ChatGPT-supported EFL § course design (ADDIE; TPACK ‡)	Taiwan	Language faculty	Case study	Practitioner (RQ1–3)
Tupper et al. (2025) [28]	ChatGPT-designed higher education field courses	UK	Academics (marine/biological; n = 3)	Case study (two designs)	Practitioner (RQ1–3)
Vazquez (2024) [21]	Responsible GenAI mini-course in aerospace engineering	Spain	Engineering faculty (n = 1); students (n = 13)	Experience report	Practitioner (RQ1–3)
Walter & Davis (2025) [22]	GenAI and instructional decision-making in reading intervention	USA	Novice teachers in university programme (n = 2)	Qualitative case study	Practitioner (RQ1–3)
Wang, L. et al. (2025) [95]	GenAI-enhanced digital storytelling system (Claude)	USA	Graduate education students *	Design case (IJDL)	Pre-service (RQ1–4)
Wang, T. et al. (2026) [96]	Collaborative ID tasks	China	Pre-service teachers (CCNU), comparison groups	Mixed methods	Pre-service (RQ1–4)
Wen & Wen (2024) [97]	Pre-service teachers' unit design with ChatGPT	USA	Pre-service teachers *	Document analysis of AI conversations	Pre-service (RQ1–4)
Wyss & Hess (2026) [98]	Simulation content: objectives, scenario, script, debriefing (INACSL)	USA (Malone)	2 nursing educators	Qualitative (design case)	Practitioner (RQ1–3)
Yang & Stefaniak (2025) [99]	Instructional designers' prioritisations for ChatGPT	USA	Practising instructional designers (n = 19)	Q methodology	Practitioner (RQ1–3)
Yang et al. (2026) [36]	Syllabus + single-session design	China	15 university instructors (3 conditions × 5)	Quantitative non-randomised	Practitioner (RQ1–3)
Yao et al. (2025) [100]	Human–machine dialogue and ID abilities	China	Teacher-education students	Mixed methods; epistemic network analysis	Pre-service (RQ1–4)
Yao et al. (2026) [35]	Multi-agent LLM course-material generation	—	Expert instructors as raters (5 CS courses)	Artefact evaluation; QM rubric	Practitioner (RQ1–3)
Yu et al. (2026) [101]	Math lesson planning	USA	44 PSTs (24 ECE + 20 ELE), Lewis University	Mixed methods	Pre-service (RQ1–4)
Zhang et al. (2025) [102]	GAI-supported ID training; HOT and problem-solving	—	Pre-service teachers * (n = 436; 3 HEIs)	Quasi-experimental with control; mixed	Pre-service (RQ1–4)
Zhao et al. (2025) [103]	EFL curriculum design through Lesson Study	China	6 university EFL teachers (Jilin)	Qualitative	Practitioner (RQ1–3)

Table 1. *Cont.*

Study	Focus	Country	Participants	Design	Set ††
Zheng & Stewart (2024) [31]	Reframing EFL § moral stories with ChatGPT: bias mitigation	Australia	Researcher-educators (n = 2)	Writing experiments (appraisal analysis)	Practitioner (RQ1–3)

* Pre-service teachers: undergraduate or postgraduate students enrolled in HE teacher-education programmes; their design activity occurs within HE coursework. † SWOT: strengths, weaknesses, opportunities, threats. ‡ TPACK: Technological Pedagogical Content Knowledge [45]. § EFL: English as a Foreign Language. †† Analytic set and research questions informed. Practitioner: practising HE educators as the studied practitioners (RQ1–RQ3; three studies additionally document self-directed professional learning under RQ4). Pre-service: pre-service teachers enrolled in HE teacher-education programmes (RQ1–RQ4).

4. Results

Corpus characteristics. The corpus comprises 78 studies spanning 2022–2026 (one from 2023, 13 from 2024, 40 from 2025, and 24 from 2026): roughly four-fifths of the evidence base was published in 2025–2026, underscoring the field's exponential growth. The practitioner strand comprises 46 studies (59%) and the pre-service strand 32 (41%). The United States remains the most represented context, but the 2025–2026 literature broadens the geographical base with contributions from South Africa, Algeria, Canada, Australia, Germany, Finland, Estonia, and the Philippines, alongside a markedly stronger East Asian presence. The tool landscape diversifies rapidly: fewer than half of the 2025–2026 studies rely on ChatGPT alone, with Claude, DeepSeek, Microsoft Copilot, Kimi, open-source models (Qwen, LLaMA), and purpose-built educational agents now documented, and several studies treating the model itself as an experimental variable through multi-model comparisons [35,62,75]. Where the original studies reported model versions, these included GPT-3.5, GPT-4, GPT-4o, GPT-4o mini, and o1 for ChatGPT and other OpenAI models, Claude 3.5 Sonnet and Claude Sonnet 4.5, Gemini Advanced, DeepSeek-R1, Microsoft 365 Copilot, Qwen 2.5-72B, and LLaMA 3.1-70B; many studies did not report a version (tools per study in Supplementary Table S3). Methodologically, the 2025–2026 studies raise the corpus profile, contributing the first controlled designs (nine comparative, quasi-experimental, or randomised controlled studies with comparison or control conditions) and the three largest samples in the corpus (n = 436; n = 147; n = 130). Table 2 summarises the corpus's descriptive characteristics. Figure 2 shows the corpus's growth and composition by publication year.

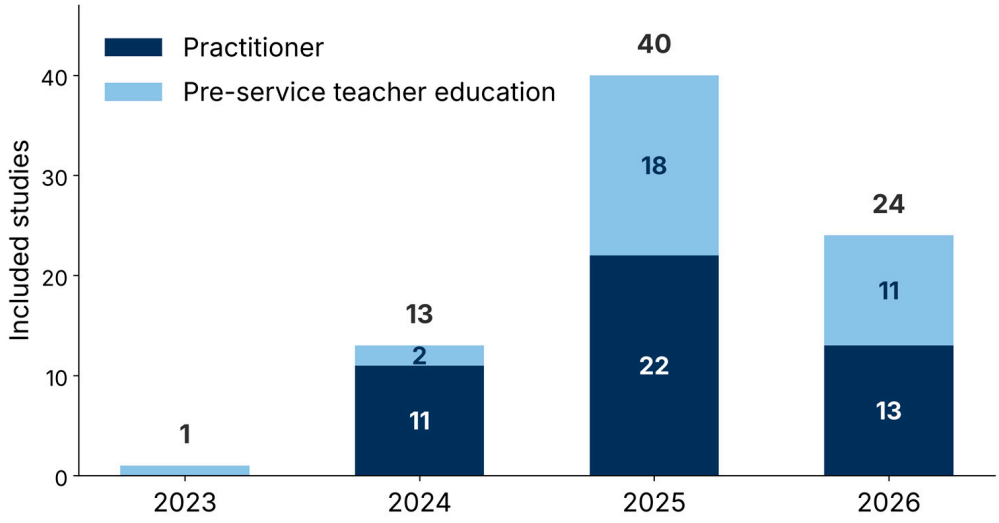


Figure 2. Included studies by publication year and analytic set (n = 78).

Table 2. Overview of the unified corpus (n = 78): descriptive characteristics.

Characteristic	Category	n (%)
Publication year	2023	1 (1.3%)
	2024	13 (16.7%)
	2025	40 (51.3%)
	2026	24 (30.8%)
Context strand	Practitioner	46 (59.0%)
	Pre-service teacher education	32 (41.0%)
Methodological design (MMAT)	Qualitative	38 (48.7%)
	Mixed methods	27 (34.6%)
	Quantitative descriptive	8 (10.3%)
	Quantitative randomised controlled	2 (2.6%)
	Quantitative non-randomised	3 (3.8%)
ADDIE phase documented †	Analysis	11 (14.1%)
	Design	74 (94.9%)
	Development	44 (56.4%)
	Implementation	7 (9.0%)
	Evaluation	10 (12.8%)
Functional AI role †	Generative	75 (96.2%)
	Scaffolding	27 (34.6%)
	Evaluative	16 (20.5%)
Human oversight †	Explicit commitment with documented practice	45 (57.7%)
	Reported at a general level only	30 (38.5%)
	Urged without described practice	3 (3.8%)
	Oversight examined empirically	21 (26.9%)

† Non-exclusive categories: a study may document several phases, roles, or oversight characteristics; percentages are of n = 78.

4.1. RQ1: Mapping AI-Supported Instructional Design Across Contexts

Figure 3 presents the phase-and-role mapping across the full corpus of 78 studies. Of the 46 practitioner studies, 20 [1,2,13,20–26,28,30–34,69,86,94,99] formed the mapping corpus of the companion paper [14]; all results below are reported for the full corpus. Four results emerge. First, AI support is structurally concentrated: Design-phase activity is near-universal (74/78, 95%) and Development appears in just over half of the studies (44/78, 56%), while Analysis (11/78, 14%), Evaluation (10/78, 13%), and Implementation (7/78, 9%) remain markedly underrepresented; no study examined AI support across the complete ADDIE cycle. Implementation, the least documented phase, is confined to practitioner studies, and only one of the seven [89] lies outside the companion mapping, indicating that the implementation gap persists as the field expands. Second, within the Design phase, AI supported four categories of design decision—learning activities, learning objectives, course structure, and assessment—with learning-objective formulation the most rigorously evaluated: only 38.3% of AI-generated learning objectives met all quality criteria without revision [25]. Third, three non-exclusive AI roles were identified: Generative (75/78, 96%), in which AI produces design content at the educator's direction; Scaffolding (27/78, 35%), in which AI structures the educator's own design reasoning, including a deliberately designed

primary role in which the AI is positioned as a critical-friend partner rather than a content generator [61]; and Evaluative (16/78, 21%), in which AI assesses artefact quality, including systems that provide structured feedback on educator-authored designs [35,38,54,76,102]. Fourth, ChatGPT remains the most frequently used tool, named in 50 of the 78 studies (64%), but the tool landscape is diversifying, with multi-model comparisons now evaluating ChatGPT alongside Claude, Gemini, and open-source models [25,35,62,75].

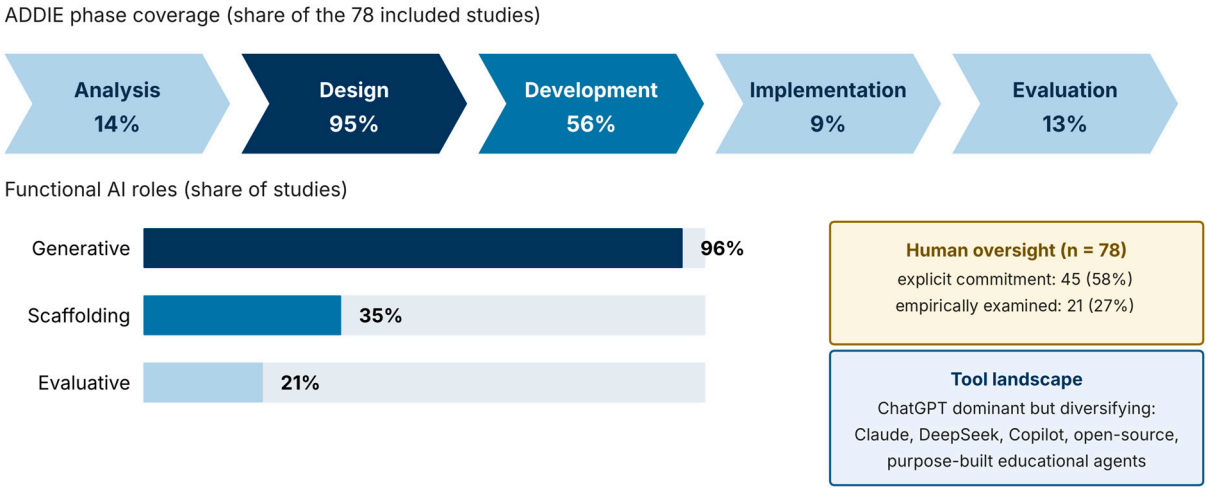


Figure 3. ADDIE phase coverage and functional AI roles across the 78 included studies, with the documented human oversight profile and the evolving tool landscape.

Disaggregating the mapping by context strand reveals a sharply profiled pre-service pattern. Design-phase activity is universal there (32/32): lesson and unit planning [27,85,97], mathematical problem posing [29], and micro-lesson design [91], with Development present in 12 of the 32 studies through material generation and adaptation [74]; Analysis and Implementation are absent, and Evaluation is documented in a single study. In the practitioner strand, AI support reaches further across the cycle (Development 32/46, Analysis 11/46, Evaluation 9/46, Implementation 7/46). The role profile also differs structurally: because the pre-service settings are pedagogically designed interventions, scaffolding is built into the AI interaction itself—structured prompting tasks, reflective journals, and checkpoint protocols [51,52,100]—so that Scaffolding appears in half of the pre-service studies (16/32) against roughly a quarter of the practitioner studies (11/46), and the AI functions simultaneously as a generative tool and as a deliberate scaffold for the developing designer. Oversight is likewise structurally distinct: it is layered, with the teacher educator supervising the pre-service teacher's supervision of the AI (Section 4.3). Table 3 summarises the cross-context comparison.

This comparison yields a finding with direct consequences for the questions that follow: within the reviewed literature, the pre-service strand functions as the laboratory for the deliberate design of AI-supported practice. Pre-service teacher education documented the most deliberately scaffolded and supervised configurations, and in developed and evaluated form; whether these configurations can transfer to, and are effective in, professional practice remains untested. Their documentation nonetheless reframes the competence (RQ2), oversight (RQ3), and development (RQ4) analyses that follow: in the reviewed literature, responsible AI-supported design has been documented predominantly where practice is institutionally supervised; the absence of documented examples in the practitioner corpus does not demonstrate their absence from professional practice itself.

Table 3. Cross-context comparison of AI-supported instructional design activity: practitioner versus pre-service corpora.

Dimension	Practitioner Corpus (n = 46)	Pre-Service Corpus (n = 32)
Studied phenomenon	Professional design practice	Competence development through design activity
Phase concentration	Design near-universal; Development in over half; Analysis, Implementation, and Evaluation sparse	Design universal (lesson and unit design); Development via material adaptation; Evaluation in a single study, other phases absent
Dominant AI role configuration	Generative dominance; scaffolding mostly emergent from iterative prompting rather than deliberately configured	Generative use embedded within deliberately designed scaffolds, including an AI positioned as a critical-friend partner [61]
Oversight structure	Predominantly single-layer, self-directed professional review; expert coaching [22] and configurable human-in-the-loop modes [35] documented in individual studies	Two-layer: teacher educator supervises the trainee's oversight of the AI
Typical evidence	Case studies, practitioner reports, self-report; comparative designs emerging	Pre–post mixed-methods designs, artefact analysis; quasi-experimental designs emerging

4.2. RQ2: Educator Competences

4.2.1. Competence Frameworks

A striking feature of the corpus is the scarcity of named competence frameworks: 59 of the 78 studies (76%) inferred competences from documented practices without theoretical mapping, and only 19 applied a named competence or teacher-knowledge framework to structure their design or analysis. TPACK and its variants [45] dominate these applications (17 studies [27,29,51–53,55,64,70,74,78,83,85,91,94,95,97,101]), applied productively as an analytical lens but stretched beyond their original scope, since TPACK was not designed for AI-specific capacities; the remaining two draw on AI literacy [60] and on the refined consensus model of pedagogical content knowledge [84]. AI-specific extensions of TPACK—Intelligent-TPACK [11] and the related AI-TPACK and GenAI-TPACK variants—appeared in five studies [51–53,64,91]; in the Aldemir studies, Intelligent-TPACK was consistently paired with AI literacy [10], with Intelligent-TPACK defining the competence domains and AI literacy specifying the developmental progression through them. Outside this count, prompt-engineering frameworks structured the AI interaction in two studies [31,34]; within it, self-regulated learning theory complemented TPACK in one [74]. Notably, none of the policy-level frameworks reviewed in Section 2.5—DigCompEdu [46], the UNESCO AI Competency Framework for Teachers [47]—were applied as an analytical framework: the UNESCO framework appears only as a source of synthesised AI literacy components [60] or as an alignment reference [52], and DigCompEdu only in passing [85]. This indicates a disconnect in both directions: empirical research proceeds largely untheorised, while the frameworks designed to guide practice remain empirically untested in the ID context. The practical consequence is that competence accounts across most of the corpus are context-specific and resistant to synthesis: without shared theoretical vocabulary, findings cannot accumulate.

4.2.2. Competence Clusters

The thematic synthesis identified seven competence clusters (Table 4). Read against the frameworks of Section 2.5, the clusters map onto Intelligent-TPACK's knowledge

domains and AI literacy's developmental progression, prompt engineering and conceptual AI literacy corresponding to the know-and-use levels, critical evaluation and pedagogical alignment to the evaluate level, and ethical awareness to the ethics strand that Intelligent-TPACK distributes across all domains [10,11], which suggests that the empirical corpus has, in effect, been rediscovering piecemeal what the integrated frameworks already articulate.

Table 4. Educator competence clusters identified through the thematic synthesis.

Competence Cluster	Key Evidence
Prompt engineering	From basic instruction formulation [27] to strategic pedagogical reasoning [22] and formalised multi-step methods [31,34].
Critical evaluation of AI outputs	Only 38.3% of AI-generated objectives met all quality criteria [25]; pre-service teachers systematically overestimate their evaluation skills [71].
Ethical awareness and bias detection	Weakest documented competence in PST studies [51,52]; WEIRD * cultural bias requires active mitigation [31].
Pedagogical–content alignment	Aligning AI content with learning goals, Bloom's levels, and disciplinary context [1,25,29]; constructive alignment as quality check [20].
Conceptual AI literacy	Understanding of LLM operations, probabilistic outputs, hallucination risk, and data governance [51,60,71].
Human oversight and validation	Near-universal normative commitment; mechanisms analysed under RQ3 (Section 4.3).
Reflective practice and self-regulation	Iterative reflection as developmental mechanism [70,85]; SRL strategies essential [74]; reflection mediates design capacity [100].

* WEIRD: Western, educated, industrialised, rich, democratic—a characterisation of the cultural values over-represented in AI training data [31]. Cluster assignment is interpretive; we derived the clusters through thematic synthesis, corroborated them across the full corpus (Section 4.2.2), and reported their prevalence in the text in indicative rather than exact terms.

Prompt engineering was the most consistently documented cluster, and the evidence reveals a developmental spectrum rather than a single skill: educators typically began with generic instruction-following prompts and progressed—through iteration, feedback, or structured training—toward contextually precise, pedagogically grounded prompting [22,27,74,97]. At its most formalised, prompt engineering appears as codified method: established techniques (zero-shot, few-shot, and chain-of-thought prompting) embedded in a question-centred debriefing model [34], and structured navigation-and-generation methods for bias mitigation [31]. This suggests that prompt engineering in ID contexts is best understood not as a technical skill but as pedagogical reasoning conducted through a new medium, the prompt as an externalised design decision.

Critical evaluation of AI outputs is empirically the best-substantiated cluster, and the evidence is sobering. Beyond the low rate of directly usable AI-generated learning objectives [25], Kerneža [71] demonstrated a systematic self-assessment gap: pre-service teachers rated their ability to interpret chatbot-generated content significantly higher than external assessors rated it. The divergence between confidence and competence recurs in Seufert et al. [91], where participants reported high confidence in technological domains while demonstrating weak ethical knowledge. This suggests that critical evaluation capacity cannot be assumed from self-report and is likely overestimated across the corpus's self-report-dominated evidence base.

Ethical awareness and bias detection emerged as the most consequential and least developed cluster. The evidence converges from multiple directions. Aldemir et al. [51] found pre-service teachers defaulted to factual accuracy checking—a cognitively accessible form of evaluation—while systematically failing to detect subtler bias: participants inter-

preted politically balanced responses as evidence of neutrality without recognising trained model bias, and identified monocultural name choices in AI-generated problems without connecting them to systematic training-data bias. Zheng and Stewart [31] demonstrated the mechanism at the material level: appraisal-theory analysis of AI-generated moral stories revealed embedded WEIRD cultural values requiring active mitigation through targeted prompt redesign. Taken together, these findings suggest that ethical competence is not primarily a knowledge deficit but a perceptual one, educators can name ethical risks they cannot detect in practice.

The framework gap shows no sign of closing as the field matures: no competence framework constructed specifically for GenAI-supported instructional design has yet appeared in the corpus; the framework applications in the most recent studies are again predominantly TPACK-based [53,55,64,78,83,95,101], with only first attempts to adapt TPACK to generative AI—an appeal to a GenAI-specific TPACK variant [64] and a design-case reconceptualisation of each TPACK domain [95], while the range of borrowed scaffolds widens further (Universal Design for Learning [56], Quality Matters [35], Learning Engineering [58], higher-order-thinking frameworks [102], teacher-agency frameworks [41]), reinforcing the field's reliance on adjacent lenses and the case for an ID-specific competence framework.

4.3. RQ3: Operationalising Human Oversight

Almost all studies (75/78, 96%) described human review of AI-generated content as part of the documented practice, and 45 (58%) did so as an explicit commitment backed by a concrete review practice, yet oversight was far more often asserted as a principle than examined as a practice. The typology of five core oversight mechanisms introduced in the companion paper for the practitioner corpus [14] is here consolidated and extended with two emerging mechanisms documented in the most recent literature across the full corpus (Table 5), ordered from proceduralised mechanisms (ethical checkpoints; rubric evaluation) through interactional mechanisms (iterative prompting; expert coaching) to attitudinal stances (professional caution). The typology draws exclusively on explicitly documented mechanisms; where oversight was endorsed but its enactment was not described, studies contribute only to the endorsement counts. The variable depth of documentation is therefore not a hidden assumption of RQ3 but part of its answer: it constitutes the oversight gap reported below.

A shift from asserted to examined oversight is beginning. Oversight has long been endorsed far more often than it has been studied; twenty-one of the 78 studies (27%) now examine oversight empirically, through criterion-based rating of AI-generated outputs by educators or expert panels [25,26,62,67,72,75], blinded rubric evaluation with inter-rater reliability [36,56] and blinded comparison of expert, teaching-assistant and GenAI feedback [38], comparison of AI and human ratings of lesson plans [37], Rasch-modelled measurement of educator-authored plans [84], analysis of how educators evaluated, revised, or rejected AI outputs [51,53,57,71,82,92,97], trace-based analysis of critical evaluation behaviour [41], comparison of design artefacts produced under an ethical-checkpoint regime [52], and a systematic comparison of human-involvement configurations in which the fullest co-pilot arrangement outperformed autonomous generation [35]. Two mechanism types extend the typology of Table 5: configurable human-in-the-loop modes [35] and dialogic critical-friend protocols [61]; accuracy-verification protocols [95,98] broaden the existing checkpoint mechanism beyond ethics to factual accuracy. Oversight nonetheless remains a stated principle rather than an examined practice in the majority of the corpus.

Table 5. Typology of human oversight mechanisms documented across the corpus.

Oversight Mechanism	Studies	Operational Form
Ethical decision checkpoints	[51,52,95,98]	Structured privacy, bias, transparency, and accountability review required before each design task; proceduralised within task design.
Structured rubric evaluation	[25,26,36,56,62,64,65,67,84]	AI outputs rated against predefined quality criteria (SMART *; Bloom's alignment); the mechanism most often accompanied by quantified quality assessments, although these rarely isolate its effect as an oversight mechanism.
Iterative prompting with feedback loops	[2,24,28,57,59,63,66,93]	Sequential refinement of prompts; each output reviewed and re-prompted until instructionally adequate; effective but time-intensive.
Collaborative expert coaching	[22]	Novice educators' AI use guided by experienced practitioners; oversight distributed across a coaching layer.
Professional cautious stance	[13,68,88,99,103]	Self-directed conscientious review shaped by institutional accountability; attitudinal rather than procedural.
Configurable human-in-the-loop modes	[35]	Graded involvement configurations (autonomous → full co-pilot); the fullest co-pilot mode outperformed autonomous generation—the first comparative examination of oversight intensity.
Dialogic critical-friend protocols	[61]	Persona-based, iterative dialogue in which the AI is explicitly positioned as a reflective scaffold rather than a generator; oversight is built into the workflow itself.

* SMART: specific, measurable, achievable, relevant, time-bound.

The typology exposes two structural findings. First, an evidential asymmetry: quantified assessments were concentrated in rubric-based evaluations of output quality [25,26,36,56], which generally did not isolate the effect of rubric use as an oversight mechanism; direct comparative evidence on oversight configurations remains sparse and heterogeneous [35], and most remaining mechanisms were described but not measured. Second, a layering phenomenon specific to teacher education: the pre-service studies add a two-layer oversight structure absent from the practitioner corpus, in which the educator-in-training's oversight of the AI is itself overseen by a teacher educator [51,52,71], an arrangement that operationalises oversight as a taught, supervised practice rather than an assumed disposition.

Only a small number of studies link a specified oversight arrangement to differences in design outputs or design quality—lesson plans produced under an ethical-checkpoint regime contained more explicit safeguards and verification steps [52], and the fullest human-in-the-loop configuration outperformed autonomous generation [35]—and the available evidence is too sparse and heterogeneous to support mechanism-level conclusions. Set against the policy landscape of Section 2.4, where human oversight is a binding regulatory obligation under the EU AI Act for AI systems classified as high-risk, a category that can encompass certain educational uses [44] and a consistent demand for international guidance [3,43], this indicates that the empirical foundations of the field's central governance concept remain remarkably thin: institutions are being asked to guarantee a practice whose mechanisms research has barely begun to specify, let alone evaluate.

4.4. RQ4: Professional Development Approaches

The studies documenting PD interventions or developmental contexts for AI-supported ID competences synthesise into six approach families (Table 6). Because the six families differ in aims, duration, modality, and implementation context, we compare them on their documented design features rather than holistically.

Table 6. Professional development approaches documented in the corpus.

PD Approach	Studies	Key Design Features
Compact embedded modules (2–4 weeks)	[51,52,102]	Ethics-threaded activities; performance-based artefacts (lesson plans, decision logs, micro-teaching); ethical checkpoints at every task; pre–post evaluation.
Semester-long microlearning	[74,88]	Thirteen-week low-cognitive-load modules; self-regulated learning strategies; multi-tool exposure.
Task-based design cycles	[29,55,61,70,85]	Iterative design tasks with reflective journals; AI problem generation; TPACK-framed micro-lessons.
Collaborative dialogue-based	[100]	Human–machine collaborative dialogue; progressive question chains; reflection as mediating mechanism.
Critical co-discovery	[60]	Educators as active agents; co-construction of AI concepts and ethics; participatory rather than transmissive.
Self-directed practitioner learning	[13,24,99]	SWOT * reflection workshops; Q-methodology viewpoint typologies; conscientious professional integration without formal training.

* SWOT: strengths, weaknesses, opportunities, threats.

The families differ substantially in duration (two weeks to a full semester), structure (embedded modules to self-directed learning), and evidential rigour (pre–post mixed-methods designs to reflective practitioner accounts).

The six approach families remain stable across the corpus, strengthening the typology's robustness. Two families gain materially: semester-long microlearning acquires a second exemplar [88], and compact embedded modules now include the largest controlled evaluation in the corpus, a four-week programme with 436 pre-service teachers and a control condition [102], which constitutes the strongest outcome evidence for any documented approach. Structured, short-format prompting training [50] emerges as a possible seventh variant, pending further exemplars.

Outcome evidence is strongest for the embedded-module approach. Aldemir et al. [52] reported statistically significant gains across all six Intelligent-TPACK domains in a three-week module, with medium-to-large effect sizes, and documented that ethical decision-making checkpoints coincided with more explicit safeguards in lesson artefacts. The companion two-week study [51] found rapid gains in knowledge and confidence but identified bias detection as the weakest emerging competence—a finding that persisted in the longer module, consistent with the possibility, not yet tested longitudinally, that ethical perceptual capacity develops more slowly than technological familiarity.

Three design features distinguish the most rigorously evidenced interventions: ethics embedded as distributed practice rather than a stand-alone session [51,52,91]; performance-based competence assessment, lesson plans, decision logs, micro-teaching—motivated directly by the demonstrated unreliability of self-report [52,71]; and prompt engineering developed as reflective, pedagogically grounded practice rather than technical instruction [22,74]. Reflection recurs as a design feature across approaches and is a plausible developmental mechanism: Yao et al. [100] found that AI-supported collaborative dialogue affected instructional design capacity indirectly, through changes in reflection ability. This

suggests, as a hypothesis requiring further testing, that PD interventions may benefit from targeting reflective capacity rather than tool proficiency as an end in itself. Conversely, persistent limitations remain: ethical knowledge remained weak despite framework-grounded, multi-method design [91]; self-directed practitioner learning produced the least transferable evidence [13,99]; and no study followed participants beyond the intervention period, leaving the durability of PD effects unknown.

5. Discussion

Taken together, the findings support a central reading: generative AI is currently used mainly to support the production of instructional-design outputs, while the evidence does not yet show that educators are systematically prepared to exercise the pedagogical judgement, critical evaluation, and human oversight that these processes require. Three arguments develop this reading. First, AI support is dominant in design and development but weakly evidenced across the rest of the instructional-design cycle (Section 5.1). Second, there is a gap between tool use and competence development: the literature seldom uses a shared framework to define or assess the competences it seeks to develop (Section 5.2). Third, researchers frequently call for human oversight but rarely evaluate it, leaving open which mechanisms actually improve design outputs or educators' decisions (Section 5.3).

5.1. Findings in Relation to Prior Literature

The research field approaches GenAI primarily as a tool for producing designs, materials, and assessments rather than as part of a complete instructional-design cycle: the mapping of Section 4.1 revealed a markedly uneven distribution of AI support, with near-universal presence in the Design phase and limited coverage elsewhere. This pattern both confirms and reframes established critiques of the AIED literature. Our findings agree with Zawacki-Richter et al. [5] and subsequent meta-reviews [6,8] that the educator's perspective remains structurally underrepresented, but they relocate the diagnosis: even within the emerging educator-facing literature, attention concentrates on the most visible, production-oriented slice of professional practice. Unlike prior reviews of AI in HE, which have organised evidence by application type or stakeholder [6,8], the ADDIE-mapped analysis makes visible a within-practice imbalance that application-level taxonomies cannot detect: it is not only that educators are under-studied, but that the parts of their design work most consequential for quality assurance—analysis and evaluation—are the least examined.

Three plausible mechanisms may account for the Design-phase concentration: tool affordance (generative models are architecturally optimised for content production, and Design and Development are the phases whose outputs are directly producible artefacts); workload salience (course preparation, spanning design and materials development; is widely regarded as among time-intensive components of educator workload, making it a natural target for efficiency-seeking adoption [1]); and evidential accessibility (Design and Development produce immediately inspectable artefacts, whereas Analysis and Evaluation involve judgements dispersed across time, actors, and data sources, and are therefore harder to study). The quality findings similarly consolidate individual results into a pattern: the low proportion of directly usable AI-generated learning objectives [25], the specificity deficits found by Doyle et al. [26], and the missing local knowledge documented by Tupper et al. [28] point to a related pattern: fluent outputs lacking sufficient contextual grounding, observed through different instruments.

5.2. Theoretical Implications

From tool to design partner. Read through the human–AI collaboration lens of Section 2.3, the review's principal findings, Design-phase concentration, Generative-role

dominance, and the oversight gap, are not three separate observations but are consistent with a single underlying pattern: generative AI appears to have been taken up in HE instructional design primarily as a productivity-enhancing tool rather than as a design partner, at least as reflected in the research literature to date. A productivity framing would explain the phase concentration (AI applied where output is produced), the role distribution (generation prioritised over scaffolding and evaluation), and the oversight gap (review treated as friction to be minimised rather than as the core professional act). The counter-model is visible in embryonic form: the scaffolding deployments [22,34] sketch a hybrid configuration in the sense of Molenaar [39], AI as a partner in, and provocation for, design thinking, with the educator's evaluative judgement at the centre of the practice rather than as its residue. The evolution the field now faces is therefore not primarily technical but configurational: whether human–AI ID practice consolidates around automation of production or augmentation of design reasoning [40].

The agency continuum and a tentative inverse pattern. The Generative–Scaffolding–Evaluative taxonomy describes a continuum in the distribution of cognitive work between educator and machine. A tentative pattern across the corpus deserves emphasis: the pedagogical value of AI-supported ID appears inversely related to the degree of cognitive work the educator delegates to the AI. The richest professional-learning outcomes appeared where AI generated the fewest and scaffolded the most [22,34]; the starkest quality problems arose where AI carried the heaviest generative load [25,26,28]. Given the heterogeneity of study designs and outcome measures across the corpus, treat this association as a hypothesis for future research rather than an established relationship. One interpretation, should the pattern be confirmed, is that the field's predominantly Generative orientation is optimised for short-term production efficiency rather than for the long-term development of educators' own design expertise. An alternative and complementary interpretation of this pattern should also be acknowledged. The quality deficits observed under extensive generative delegation may be attributable not only to the delegation configuration itself, but also to the limitations of current general-purpose systems, which have not been trained for the specific demands of instructional design. Human–AI design collaboration therefore implicates a third, less visible actor: the developers and trainers of the underlying systems, whose design and training decisions are embedded in the tool with which the educator interacts. The implications of this third actor are considered further in Section 7.

A de facto linearisation of ADDIE. When AI accelerates Design and Development while Analysis and Evaluation remain manual or are bypassed, design economics may shift: producing new content becomes cheaper than evaluating whether existing content works. Moreover, when evaluation does occur, its outputs tend to feed new content production rather than the systematic redesign of existing designs, reinforcing the linear drift. The unexamined feedback loop from Evaluation back to Analysis is therefore not merely a research gap; it may signal an emerging structural change in design practice that would invert the quality logic for which systematic ID frameworks were constructed [7]. The framework gap compounds the risk, and its meaning should be stated precisely: the issue is not that educators lack competences, but that the literature seldom uses a shared framework to define or assess the competences it seeks to develop. With three-quarters of studies (76%) operating without named competence frameworks, the field lacks the shared vocabulary through which such structural drift could even be collectively recognised, a maturity problem directly analogous to the pre-TPACK fragmentation of educational technology research that Mishra and Koehler's framework resolved [45].

Ethics as perceptual expertise. The convergent ethics findings, defaulting to accuracy checking while overlooking systemic bias [51], weak ethical knowledge despite high

technological confidence [91], self-assessment diverging from external assessment [71], permit a more precise diagnosis than "insufficient awareness." They indicate a perceptual rather than declarative deficit: detecting bias in AI-generated instructional content is, for the educator, a trained pattern-recognition capacity built through repeated, feedback-rich encounters with instances, not through general declarative teaching about bias—an interpretation advanced here as a hypothesis generated by the synthesis rather than an established empirical diagnosis. This framing explains why the embedded-checkpoint intervention [52] succeeded with awareness-raising plateaus, making the ethical check an explicit procedural step obliges educators to perform the act of detection repeatedly in context, until it becomes fluent, and why bias detection remained the weakest competence even after intervention [51]: plausibly because perceptual expertise develops more slowly than declarative knowledge or tool fluency—a conjecture that awaits longitudinal testing.

5.3. Policy and Practical Implications

Perhaps the clearest contribution of this synthesis is showing that the literature endorses human oversight as a principle while offering limited evidence on which mechanisms actually improve design outputs or educators' decisions. For universities and institutional governance, this gap carries regulatory weight. The EU AI Act makes human oversight a binding obligation for high-risk educational AI uses [44], and international guidance converges on the same expectation [3,42,43]—yet this review finds oversight stated as an explicit commitment in 58% of studies (45/78) and examined empirically in only 21 (27%). Policies that mandate "human oversight" without specifying mechanisms risk compliance in name only. Quality-assurance frameworks need operational content: review protocols, quality criteria, documented checkpoints. Emerging frameworks that translate ethical principles into role-specific operational guidance for HE stakeholders [48] show the direction needed, and the oversight typology of five core and two emerging mechanisms (Table 5) offers institutions an evidence-informed, descriptive starting point for constructing such protocols, pending empirical validation of each mechanism.

For instructional designers and faculty, the finding that only 38.3% of AI-generated learning objectives met all quality criteria without revision [25] is a reminder that AI-generated design artefacts are drafts requiring systematic review, and the role taxonomy offers a reflective instrument: is the AI generating my design, or scaffolding my design thinking?

For professional development providers and designers, the features that distinguish the most rigorously evidenced approaches—embedded ethics checkpoints, performance-based assessment, and reflective prompt engineering practised on authentic design tasks—offer a concrete template (Table 6), each with a theoretical rationale: ethics as perceptual expertise requiring situated practice, self-report as demonstrably unreliable, and reflection as a plausible developmental mechanism [100].

For policy-makers, the oversight gap indicates that mandating oversight is insufficient without specifying mechanisms: policy instruments should reference concrete, auditable practices of the kind catalogued in Table 5 rather than generic endorsements. For distance and online institutions specifically, the scarcity of distance HE contexts in the corpus—few studies explicitly examine distance-education institutions, and Kumar et al. [13] provides the clearest case—should be read as a warning: the sector in which design decisions carry the greatest scale currently has the thinnest evidence base for AI-supported design practice.

5.4. Limitations

Read the findings in light of some limitations. Regarding the evidence base: the English-language search under-represents research published in other languages, although

several included studies originate from non-Anglophone contexts; the corpus is methodologically heterogeneous, practitioner reports, small-sample case studies, and mixed-methods designs, making quantitative synthesis inappropriate; self-report methodologies dominate, so the demonstrated divergence between self-assessed and externally assessed competences [71] applies as a general caution; and the rapid evolution of AI tools means that findings on tool limitations may not reflect current capabilities. Regarding the review process: screening and coding were performed by the first author, with borderline cases and ambiguities resolved through consensus with the second author; coding reliability was not quantified through independent double-coding, so the interpretive categories, particularly the AI roles and oversight levels, carry a risk of single-coder bias, mitigated but not eliminated by the second author's full review of the extraction file and by the study-level reporting in Supplementary Table S3. We appraised methodological quality with MMAT 2018; the appraisal informed the confidence assessment of the findings rather than study exclusion, so lower-quality studies contribute to the synthesis with correspondingly lower evidential weight. In addition, study designs and evidence types are reported transparently for every study (Table 1), and evidential strength is weighed explicitly throughout the synthesis. Regarding retrieval and currency: we could not retrieve 22 reports despite institutional access, as documented in the flow diagram, and the 2026 literature is necessarily incomplete, as indexing of late-2026 publications continues.

6. Synthesis: An Integrated View of the Findings

The four research questions were analysed separately, but their findings are structurally interdependent, and this section assembles them into an integrative synthesis (Figure 4). A confidence assessment for the principal synthesised findings, integrating the MMAT appraisal with the number, consistency, and directness of the supporting studies, is provided in the Supplementary Materials (Table S4). At its top sits the ADDIE design cycle—the professional process within which all AI support is situated, and whose iterative quality logic defines what complete integration would mean. Three interdependent pillars sustain this cycle. The AI roles pillar (RQ1; extending the companion paper [14]) describes what the technology contributes and where generative production concentrated in Design and Development, scaffolding of design reasoning, and emergent evaluation of artefacts. The educator competences pillar (RQ2) describes what each role requires to function: prompt engineering for the generative role, critical evaluation and pedagogical alignment for output adoption, ethical perception for bias detection, and reflective self-regulation across all of them. The human oversight pillar (RQ3) describes the mechanisms through which competence is exercised over AI contributions, spanning proceduralised, interactional, and attitudinal forms. Professional development (RQ4) underpins all three pillars by building competences and embedding oversight mechanisms through ethics, performance-based assessment, and reflective practice, feeding capacity back into the design cycle. Across the corpus, the principal findings hold consistently: the Design-phase concentration and the framework gap persist in the most recent literature; the oversight gap persists, although empirically examined mechanisms now form a measurable minority (21/78); and the field is moving from generic ChatGPT applications towards a diversified tool landscape and more complex forms of human–AI collaboration.

The synthesis makes the review's central claims visible as relations rather than isolated findings. The productivity-tool pattern of Section 5.2 appears here as an imbalance among pillars: a hypertrophied generative role, competences developed unevenly, and an oversight pillar rich in commitment but poor in mechanism. Correspondingly, the synthesis specifies what a mature configuration would require: AI roles distributed across the full cycle, competences theorised through shared frameworks, oversight operationalised in

documented mechanisms, and professional development designed to sustain all three. This synthesis consolidates the review's empirical contribution; it is deliberately descriptive rather than prescriptive. A complementary theoretical contribution—a conceptual framework of AI-augmented pedagogical decision-making—is being developed in subsequent work within the AI-CODE project, for which the present synthesis provides the empirical foundation.

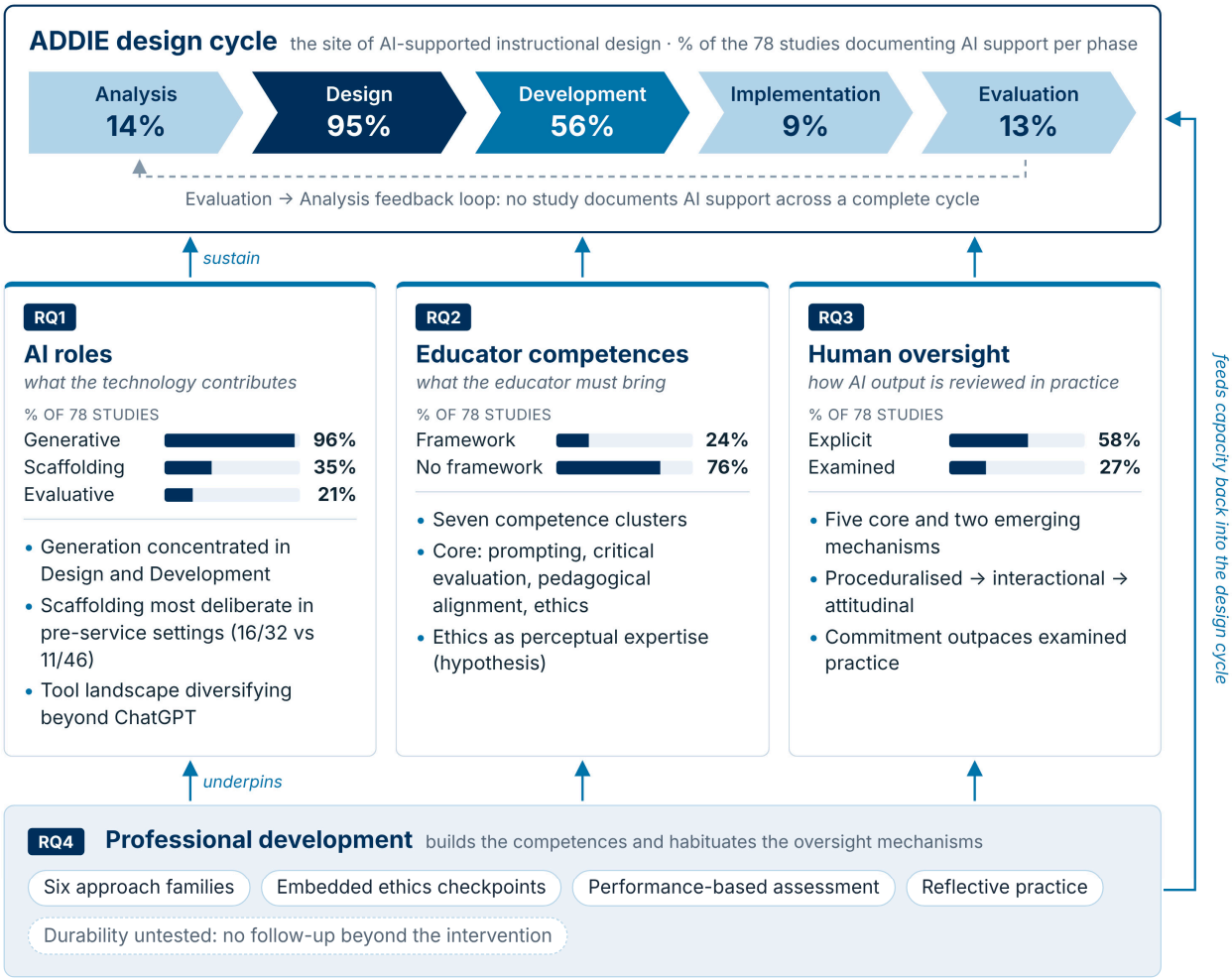


Figure 4. Integrative synthesis of the review findings: the ADDIE cycle as the site of AI-supported instructional design, sustained by three interdependent pillars (AI roles, educator competences, human oversight mechanisms) and underpinned by professional development. Solid vertical arrows indicate that the three pillars sustain design practice within the ADDIE cycle and that professional development underpins the pillars; the side arrow indicates that professional development feeds capacity back into the design cycle; the dashed arrow marks the Evaluation-to-Analysis feedback loop, which no included study documents across a complete cycle. Percentages are shares of the 78 included studies.

7. Future Research

The review's findings define a concrete research agenda, organised here from the design process outward. First, complete-cycle studies: no study examined AI support across a full ADDIE iteration; longitudinal designs following a course from analysis through evaluation and redesign would test whether the de facto linearisation identified here materialises in practice. Second, the underrepresented phases: empirical work on AI-supported learner and context analysis, and on formative course evaluation and redesign, where the corpus is the thinnest. Third, oversight mechanisms: observational and artefact-based

research on how educators actually exercise oversight, and with what effect on design quality, moving the field's central governance concept from commitment to evidence. Fourth, the scaffolding role: dedicated investigation of AI as a structure for design reasoning, including the tentative inverse relationship between cognitive delegation and professional learning proposed in Section 5.2. Fifth, comparative designs: systematic comparisons across disciplines, across AI tools, and between professional instructional designers and faculty designers, whose competence profiles and oversight practices plausibly differ [13,99]. Sixth, distance and online contexts: the sector where design decisions carry the greatest scale is nearly absent from the corpus and should be a priority setting for all of the above. Seventh, competence research: framework-grounded studies enabling cross-study accumulation—including empirical application of DigCompEdu [46] and the UNESCO AI Competency Framework [47] to ID practice, alongside focused work on prompt engineering as pedagogical competence and on ethical perception, with performance-based rather than self-report measurement [71]. Finally, longitudinal PD studies: no intervention study followed participants beyond the intervention period; the durability of competence gains, particularly the slow-developing ethical perceptual capacities, is an open question of direct institutional relevance. A further direction concerns the third, largely invisible party in human–AI instructional design: the developers and trainers of the GenAI systems themselves. Educator use generates precisely the interaction data on which such systems are refined, suggesting a human–AI co-evolution agenda [104] in which educator-informed feedback, participatory co-design [105], fine-tuning, and evaluation shape the next generation of design-support tools—and in which the division of responsibility among educators, institutions, and system providers is made explicit.

8. Conclusions

This review shows that generative AI is becoming embedded in higher-education instructional design, particularly in the design and development of learning objectives, materials, and assessments. The evidence base, however, remains uneven: most studies apply no explicit competence framework—Intelligent-TPACK, used alongside AI literacy in a small number of studies, offers a theoretically promising integrated basis, although its empirical use in instructional-design practice remains limited—and human oversight is more often recommended than empirically evaluated. Pre-service teacher education documented the most deliberately scaffolded and supervised configurations; whether these configurations transfer to, and are effective in, professional practice remains untested.

The findings suggest that professional development should move beyond tool-oriented training and support educators in evaluating AI-generated outputs, making pedagogical decisions, and documenting accountable human oversight throughout the design process. Future research should test whether such embedded support improves the quality, inclusiveness, and pedagogical alignment of educator-produced instructional designs. The integrative synthesis of Section 6 assembles the findings into a coherent account of AI-supported instructional design—one in which the educator's design expertise, rather than the technology's generative fluency, remains the organising centre of the practice.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/computers15110745/s1, Table S1: Search execution details per database; Table S2: Per-study methodological quality appraisal (MMAT 2018); Table S3: Study-level coding matrix of the 78 included studies; Table S4: Confidence assessment of the principal synthesised findings; PRISMA 2020 checklist. References [1,2,13,14,20–38,41,49–103] are cited in the Supplementary Materials.

Author Contributions: Conceptualization, N.D.S. and A.D.K.; methodology, N.D.S.; software, N.D.S.; validation, N.D.S. and A.D.K.; formal analysis, N.D.S.; investigation, N.D.S.; resources, N.D.S.; data curation, N.D.S.; writing—original draft preparation, N.D.S.; writing—review and editing, N.D.S. and A.D.K.; visualization, N.D.S.; supervision, A.D.K.; project administration, N.D.S.; funding acquisition, A.D.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Hellenic Open University, grant number 80700, within the AI-CODE project (Artificial Intelligence Competences for Online and Distance Education).

Data Availability Statement: The original contributions presented in this study are included in the article and its Supplementary Materials (study characteristics in Table 1; per-study quality appraisal and coding in Supplementary Tables S2 and S3).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Tran, T.M.; Bakajic, M.; Pullman, M. Teacher's pet or rebel? Practitioners' perspectives on the impacts of ChatGPT on course design. *High. Educ.* **2025**, *90*, 777–800. [CrossRef]
2. Conklin, S.; Dorgan, T.; Barreto, D. Is AI the new course creator? *Discov. Educ.* **2024**, *3*, 285. [CrossRef]
3. UNESCO. *Guidance for Generative AI in Education and Research*; UNESCO: Paris, France, 2023. [CrossRef]
4. Holmes, W.; Tuomi, I. State of the art and practice in AI in education. *Eur. J. Educ.* **2022**, *57*, 542–570. [CrossRef]
5. Zawacki-Richter, O.; Marín, V.I.; Bond, M.; Gouverneur, F. Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *Int. J. Educ. Technol. High. Educ.* **2019**, *16*, 39. [CrossRef]
6. Crompton, H.; Burke, D. Artificial intelligence in higher education: The state of the field. *Int. J. Educ. Technol. High. Educ.* **2023**, *20*, 22. [CrossRef]
7. Branch, R.M. *Instructional Design: The ADDIE Approach*; Springer: Berlin/Heidelberg, Germany, 2009. [CrossRef]
8. Bond, M.; Khosravi, H.; De Laat, M.; Bergdahl, N.; Negrea, V.; Oxley, E.; Pham, P.; Chong, S.W.; Siemens, G. A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *Int. J. Educ. Technol. High. Educ.* **2024**, *21*, 4. [CrossRef]
9. Lo, C.K. What is the impact of ChatGPT on education? A rapid review of the literature. *Educ. Sci.* **2023**, *13*, 410. [CrossRef]
10. Ng, D.T.K.; Leung, J.K.L.; Chu, S.K.W.; Qiao, M.S. Conceptualizing AI literacy: An exploratory review. *Comput. Educ. Artif. Intell.* **2021**, *2*, 100041. [CrossRef]
11. Celik, I. Towards Intelligent-TPACK: An empirical study of teachers' professional knowledge to ethically integrate artificial intelligence. *Comput. Hum. Behav.* **2023**, *141*, 107620. [CrossRef]
12. Conole, G. *Designing for Learning in an Open World*; Springer: Berlin/Heidelberg, Germany, 2013. [CrossRef]
13. Kumar, S.; Gunn, A.; Rose, R.; Pollard, R.; Johnson, M.; Ritzhaupt, A.D. The role of instructional designers in the integration of generative artificial intelligence in online and blended learning in higher education. *Online Learn.* **2024**, *28*, 207–231. [CrossRef]
14. Spyropoulou, N.; Kameas, A. Beyond lesson planning: Generative AI support across the full instructional design cycle in higher education. In *Proceedings of the IMCL2026 Conference*; Lecture Notes in Computer Science, in press; Springer: Berlin/Heidelberg, Germany, 2026.
15. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [CrossRef] [PubMed]
16. Wiggins, G.; McTighe, J. *Understanding by Design*, 2nd ed.; Association for Supervision and Curriculum Development (ASCD): Arlington, VA, USA, 2005.
17. Biggs, J. Enhancing teaching through constructive alignment. *High. Educ.* **1996**, *32*, 347–364. [CrossRef]
18. CAST. *Universal Design for Learning Guidelines Version 2.2*; CAST: Lynnfield, MA, USA, 2018. Available online: http://udlguidelines.cast.org (accessed on 20 October 2026).
19. Razzouk, R.; Shute, V. What is design thinking and why is it important? *Rev. Educ. Res.* **2012**, *82*, 330–348. [CrossRef]
20. Squalli Houssaini, M.; Aboutajeddine, A.; Toughrai, I.; Ibrahimi, A. Development of a design course for medical curriculum: Using design thinking as an instructional design method empowered by constructive alignment and generative AI. *Think. Ski. Creat.* **2024**, *52*, 101491. [CrossRef]
21. Vazquez, R. Experiences and insights from a mini-course on responsible generative AI use in aerospace engineering. *IFAC-PapersOnLine* **2024**, *58*, 35–40. [CrossRef]
22. Walter, A.R.; Davis, D.S. Not magic but manageable: Generative AI and instructional decision-making in structured reading intervention. *Read. Teach.* **2025**, *79*, e70036. [CrossRef]

23. Choi, G.W.; Kim, S.H.; Lee, D.; Moon, J. Utilizing generative AI for instructional design: Exploring strengths, weaknesses, opportunities, and threats. *TechTrends* **2024**, *68*, 832–844. [CrossRef]
24. Kumar, J.A.; Zhuang, M.; Thomas, S. ChatGPT for natural sciences course design: Insights from a strengths, weaknesses, opportunities, and threats analysis. *Nat. Sci. Educ.* **2024**, *53*, e70003. [CrossRef]
25. Crowther, G.J.; Funk, M.D.; Hennessey, K.M.; Lawrence, M.M. Frontier model chatbots can help instructors create, improve, and use learning objectives. *Adv. Physiol. Educ.* **2025**, *49*, 653–659. [CrossRef] [PubMed]
26. Doyle, A.; Sridhar, P.; Agarwal, A.; Savelka, J.; Sakr, M. A comparative study of AI-generated and human-crafted learning objectives in computing education. *J. Comput. Assist. Learn.* **2025**, *41*, e13092. [CrossRef]
27. Kerr, R.C.; Kim, H. From prompts to plans: A case study of pre-service EFL teachers' use of generative AI for lesson planning. *Engl. Teach.* **2025**, *80*, 95–118. [CrossRef]
28. Tupper, M.; Hendy, I.W.; Shipway, J.R. Field courses for dummies: To what extent can ChatGPT design a higher education field course? *Innov. Educ. Teach. Int.* **2025**, *62*, 512–526. [CrossRef]
29. Segal, R.; Biton, Y. The contribution that utilizing generative AI for problem posing makes to pre-service high school mathematics teachers' TPACK. *Int. J. Educ. Math. Sci. Technol.* **2024**, *12*, 1559–1582. [CrossRef]
30. Rodríguez Jerez, S.A.; Leyva Casas, J.S.; Rueda Osorio, M. Integration of generative artificial intelligence and 3D immersive environments in competency-based higher education. *Educ. Process Int. J.* **2025**, *19*, e2025544. [CrossRef]
31. Zheng, Y.D.; Stewart, N. Improving EFL students' cultural awareness: Reframing moral dilemmatic stories with ChatGPT. *Comput. Educ. Artif. Intell.* **2024**, *6*, 100223. [CrossRef]
32. Kaczorowski, T.; Stockman, A.; Hashey, A.; Kaczorowski, J. Early adopters: Navigating AI integration in special education teacher preparation. *J. Spec. Educ. Prep.* **2024**, *4*, 18–29. [CrossRef]
33. Daniels, M.; Kelly, É.; Flynn, S.; Kelly, J. Advancing project leadership education through AI-enhanced game-based learning. *Proj. Leadersh. Soc.* **2025**, *6*, 100189. [CrossRef]
34. Jang, A.; Oh, M.; Song, M.O. Question-centered prompt engineering for nursing simulation debriefing: Model development and validation. *Nurse Educ. Pract.* **2025**, *88*, 104541. [CrossRef] [PubMed]
35. Yao, H.; Xu, W.; Turnau, J.; Kellam, N.; Wei, H. Instructional agents: Reducing teaching faculty workload through multi-agent instructional design. In *Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics*; Association for Computational Linguistics: Kerrville, TX, USA, 2026; Volume 1, pp. 4087–4109. [CrossRef]
36. Yang, X.; Li, Z.; Zhao, H.; Li, J. Teaching AI assistant: A retrieval-augmented intelligent system for university teaching design. In *Proceedings of EDCS 2026*; Association for Computing Machinery: New York, NY, USA, 2026. [CrossRef]
37. Hauk, D.; Soujon, N. How reliable are large language models in analyzing the quality of written lesson plans? *Comput. Educ. Artif. Intell.* **2026**, *10*, 100538. [CrossRef]
38. Jiang, Y.; Feng, X.-M.; Liu, Y.; Wang, Y.; Xie, L.; Bao, L. Generative AI for feedback and collaborative knowledge construction in preservice physics teacher education. *Phys. Rev. Phys. Educ. Res.* **2026**, *22*, 010116. [CrossRef]
39. Molenaar, I. Towards hybrid human-AI learning technologies. *Eur. J. Educ.* **2022**, *57*, 632–645. [CrossRef]
40. Shneiderman, B. *Human-Centered AI*; Oxford University Press: Oxford, UK, 2022. [CrossRef]
41. Sha, L.; Liang, Z.; Zhao, L.; Li, Y.; Fernández-Nieto, G.M.; Tsai, Y.-S.; Swiecki, Z.; Chen, G.; Gašević, D. How do critical evaluations of generative AI outputs shape teacher agency? *Int. J. Human–Computer Interact.* **2026**, *Advance online publication*. [CrossRef]
42. OECD. *OECD Digital Education Outlook 2023: Towards an Effective Digital Education Ecosystem*; OECD Publishing: Paris, France, 2023. [CrossRef]
43. European Commission. *Ethical Guidelines on the Use of Artificial Intelligence (AI) and Data in Teaching and Learning for Educators*; Publications Office of the European Union: Luxembourg, 2022. [CrossRef] [PubMed]
44. European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Off. J. Eur. Union* **2024**, *L 1689*.
45. Mishra, P.; Koehler, M.J. Technological pedagogical content knowledge: A framework for teacher knowledge. *Teach. Coll. Rec.* **2006**, *108*, 1017–1054. [CrossRef]
46. Redecker, C. *European Framework for the Digital Competence of Educators: DigCompEdu*; Punie, Y., Ed.; Publications Office of the European Union: Luxembourg, 2017. [CrossRef]
47. UNESCO. *AI Competency Framework for Teachers*; UNESCO: Paris, France, 2024. [CrossRef]
48. Triantafyllopoulos, L.; Takouli, V.; Kalles, D.; Kameas, A.; Antonaci, A. From theory to practice: ADMIT's framework for ethical GenAI in higher education. In *Artificial Intelligence Applications and Innovations (AIAI 2026)*; IFIP Advances in Information and Communication Technology; Maglogiannis, I., Iliadis, L., Eds.; Springer: Berlin/Heidelberg, Germany, 2026; Volume 795, pp. 299–312. [CrossRef]
49. Aguş, F.S.; Mutlu, M.E. Integrating generative AI for comprehensive ELT material development. *Innov. Lang. Learn. Teach.* **2026**, *Advance online publication*. [CrossRef]

50. Akpinar, Y.; Kutbay, E. Regulatory effects of generative AI on self-regulated learning and AI literacy: A mixed-methods study. *Turk. Online J. Distance Educ.* **2026**, *27*, 1–24. [CrossRef]
51. Aldemir, T.; Bicer, A.; Kilinc, S.; Moon, J.; Kwok, M. Exploring emergent AI-TPACK competencies in a two-week AI literacy module for preservice teachers. *Teach. Teach. Educ.* **2025**, *168*, 105231. [CrossRef]
52. Aldemir, T.; Kilinc, S.; Bicer, A.; Grant, P.; Davis, T.; Sweany, N.W. Intelligent-TPACK in practice: Design and evidence from a three-week teacher preparation module. *Comput. Educ. Open* **2025**, *9*, 100306. [CrossRef]
53. Aydin Gunbatar, S.; Tezcan Sirin, G.; Ilkyaz, O.C.; Mutlu, Y. Exploring the AI interaction profiles of participants with different levels of teaching experience for lesson planning. *Chem. Educ. Res. Pract.* **2025**, *26*, 977–995. [CrossRef]
54. Cai, H.; Han, B.; Sun, J.; Li, X.; Wong, L.-H. Harnessing AI for teacher education: ChatGPT-supported lesson plan critiques. *Internet High. Educ.* **2025**, *67*, 101022. [CrossRef]
55. Chen, Q.; Osman, K. Changes in pre-service physics teachers' TPACK and collaborative problem solving associated with an AI-supported CTD-PBL module: A quasi-experimental study. *Information* **2026**, *17*, 688. [CrossRef]
56. Chick, J.C.; Morello, L.T. Rapid AI-assisted instructional design: Using agentic LLM tools to develop UDL-aligned curricula for student veterans and multilingual learners. *Appl. Sci.* **2026**, *16*, 3871. [CrossRef]
57. Choi, Y. Integrating ChatGPT into the design of 5E-based Earth science lessons. *Educ. Sci.* **2025**, *15*, 815. [CrossRef]
58. da Silva, R.L. Capturing elusive technology: Designing a course on AI for learning and development practitioners. *Int. J. Des. Learn.* **2025**, *16*, 124–134. [CrossRef]
59. Davis, J.T. Utilization of ChatGPT to create materials for a flipped classroom. *Int. J. Des. Learn.* **2025**, *16*, 236–243. [CrossRef]
60. Dilek, M.; Baran, E.; Aleman, E. AI literacy in teacher education: Empowering educators through critical co-discovery. *J. Teach. Educ.* **2025**, *76*, 288–302. [CrossRef]
61. Faith, L.; Zaugg, T.; Stolys, N.; Szabo, M.; Haghi, F.; Badlis, C.; Olmedo, S.L. Persona, Break Glass, Name Plan, Jam (PBNJ): A new AI workflow for planning and problem solving. *AI* **2025**, *6*, 310. [CrossRef]
62. Farrag, A.N.; El-Zeiny, A.; Ali, A.M. Evaluating large language models for pharmacotherapy simulations: A mixed-methods study. *npj Digit. Med.* **2026**, *9*, 355. [CrossRef] [PubMed]
63. Fernández-Sánchez, A.; Lorenzo-Castiñeiras, J.J.; Sánchez-Bello, A. Navigating the future of pedagogy: The integration of AI tools in developing educational assessment rubrics. *Eur. J. Educ.* **2025**, *60*, e12826. [CrossRef]
64. Fu, H. The impact of GenAI-assisted instructional design on the teaching ability of pre-service physical education teachers. *Sci. Rep.* **2026**, *16*, 24629. [CrossRef] [PubMed]
65. Ghouali, K. Empowering EFL learning: Leveraging ChatGPT for lesson planning and activity generation in the EFL classroom. *Onomázein* **2025**, *NE16*, 31–48. [CrossRef]
66. Gold, L.A.; Winn, V.; Arnold, J.M. Exploring the impact of generative AI on curriculum design and instruction: A study with preservice teachers. *J. Teach. Learn. Technol.* **2026**, *14*, 35–47. [CrossRef]
67. Gültekin, A.; Diri, S.; Özcan, M. An LLM-assisted decision-support system for semantic analysis of Bloom's-aligned learning outcomes. *Inf. Process. Manag.* **2026**, *63*, 104911. [CrossRef]
68. Guo, X.; Chen, H. Incorporating generative AI in foreign language teaching preparation: An empirical investigation into the effects of AI-assisted lesson planning. *Int. J. Adv. Comput. Sci. Appl.* **2026**, *17*. [CrossRef]
69. Husain, A.J.A. Potentials of ChatGPT in computer programming: Insights from programming instructors. *J. Inf. Technol. Educ. Res.* **2024**, *23*, 002. [CrossRef] [PubMed]
70. Jou, M.; Kuo, T.-H.; Chiang, Y.-C.; Hao, Y.; Huang, C.-C. Integrating generative AI in teacher education: A qualitative exploration of TPACK growth and critical reflection. *Turk. Online J. Educ. Technol.* **2025**, *24*, 54–59.
71. Kerneža, M. Fundamental and basic cognitive skills required for teachers to effectively use chatbots in education. In *Proceedings of the 5th International Scientific Conference BalticSTE2023*; Scientia Socialis: Šiauliai, Lithuania, 2023; pp. 99–110. [CrossRef]
72. Kessner, T.M.; Kaka, S.J. Can ChatGPT make quality social studies lesson plans? The answer surprised us. *Soc. Educ.* **2025**, *89*, 364–368. [CrossRef]
73. King, N.; Yan, J. Designing for AI literacy: A modular, GenAI-integrated course for interdisciplinary graduate students in education. *Int. J. Des. Learn.* **2025**, *16*, 289–305. [CrossRef]
74. Kohnke, L.; Zou, D.; Xie, H. Microlearning and generative AI for pre-service teacher education: A qualitative case study. *Educ. Inf. Technol.* **2025**, *30*, 21221–21248. [CrossRef]
75. Leinonen, J.; Zhang, L.; Hellas, A. AI-generated slides: Are they good? Can students tell? In *Proceedings of the Western Canadian Conference on Computing Education (WCCCE 2026)*; Association for Computing Machinery: New York, NY, USA, 2026. [CrossRef]
76. Li, H. Effects of GAI-supported scenario inquiry-based learning in multiplayer VR on pre-service teachers' instructional design skills. *Educ. Technol. Res. Dev.* **2026**, *Advance online publication*. [CrossRef]
77. Li, T.; Zhan, Z.; Ji, Y.; Li, T. Exploring human and AI collaboration in inclusive STEM teacher training: A synergistic approach based on self-determination theory. *Internet High. Educ.* **2025**, *65*, 101003. [CrossRef]

78. Lin, Y.; Jiang, Y.; Lei, J. Generative AI and teacher professional identity development: A case study on L2 Chinese reading materials development. *Comput. Assist. Lang. Learn.* **2025**, *Advance online publication*. [CrossRef]
79. Liu, C.-C.; Wang, D.; Gu, X.; Hwang, G.-J.; Tu, Y.-F.; Wang, Y. Facilitating pre-service teachers' instructional design and higher-order thinking with generative AI. *J. Res. Technol. Educ.* **2026**, *58*, 880–905. [CrossRef]
80. Lorenz, B.; Kikkas, K. Innovating IT ethics education: The role of AI in course development. In *HCI International 2025—Posters*; Lecture Notes in Computer Science; Smith, B.K., Borge, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2025. [CrossRef]
81. McDonald, R.; Indiogine, S.E.P.; Lachaud, N.; Lu, W.; Khokhar, M.A. Perceptions on the effectiveness of using generative AI and voice cloning to aid in the development of online course materials. In *Proceedings of the 2025 ASEE Annual Conference & Exposition*; American Society for Engineering Education: Washington, DC, USA, 2025. [CrossRef]
82. Moorhouse, B.L.; Ho, T.Y.; Wu, C.; Wan, Y. Pre-service language teachers' task-specific large language model prompting practices. *RELC J.* **2025**, *Advance online publication*. [CrossRef]
83. Moundridou, M.; Matzakos, N. Generative AI in an educational technology course for pre-service mechanical engineering educators: A case study. *Int. J. Inf. Educ. Technol.* **2025**, *15*, 867–881. [CrossRef]
84. Ndlovu, B.P.; Khoza, H.C.; Sibanda, D. Pre-service science teachers' adaptive pedagogical reasoning in AI-supported lesson plans: A case with chemical equilibrium. *Front. Educ.* **2026**, *11*, 1811975. [CrossRef]
85. Oh, S. Integration of MATH41 and generative AI in pre-service mathematics teacher education: An empirical study on lesson design competency. *IEEE Access* **2025**, *13*, 45671–45684. [CrossRef]
86. Okulu, H.Z.; Muslu, N. Designing a course for pre-service science teachers using ChatGPT: What ChatGPT brings to the table. *Interact. Learn. Environ.* **2024**, *32*, 7450–7467. [CrossRef]
87. Panke, S. How can (A)I research this? An autoethnographic exploration of generative AI in research, teaching, and design. *J. Teach. Educ.* **2025**, *76*, 230–244. [CrossRef]
88. Park, E.; Bodek, M.; Valcin, J. Prompted to explore: A microlearning model for faculty development in generative AI. *J. Teach. Learn. Technol.* **2026**, *14*, 3–19. [CrossRef]
89. Santiago, C.M. 'Generative AI made me do this': ChatGPT-assisted collaborative action research in science higher education. *Educ. Action Res.* **2026**, *34*, 634–651. [CrossRef]
90. Schilhabel, S.A. Redesigning publisher slide decks with ChatGPT: A design science approach. *Issues Inf. Syst.* **2025**, *26*, 244–257. [CrossRef]
91. Seufert, S.; Hartmann, P.; Spirgi, L. Fostering Intelligent-TPACK through AI-assistance: A multi-method study in pre-service teacher education. *Comput. Educ. Open* **2025**, *9*, 100314. [CrossRef]
92. Sun, Y.; Huang, Y. Exploring cognitive engagement in guided ChatGPT-supported collaborative lesson plan refinement. *Educ. Technol. Res. Dev.* **2026**, *Advance online publication*. [CrossRef]
93. Suresh, D.D.; Dilek, M.; Karabulut-Ilgu, A.; Baran, E.; Kimber, M. Generative AI in learning design practice: Building an online course for biomedical sciences research programs. *Int. J. Des. Learn.* **2025**, *16*, 244–265. [CrossRef] [PubMed]
94. Tseng, Y.-C.; Lin, Y.-H. Enhancing English as a foreign language (EFL) learners' writing with ChatGPT: A university-level course design. *Electron. J. E-Learn.* **2024**, *21*, 78–97. [CrossRef]
95. Wang, L.; Wang, Y.; Lewis, T.D.; Johnson, T. A GenAI-enhanced digital storytelling instructional system for civic education: Empowering student voices. *Int. J. Des. Learn.* **2025**, *16*, 205–221. [CrossRef]
96. Wang, T.; Long, T.; Mei, A.; Li, J.; Liu, Z. Fusing Six-Hat thinking with AI: Effects on pre-service teachers' lesson design. *Think. Ski. Creat.* **2026**, *59*, 102025. [CrossRef]
97. Wen, W.; Wen, H. Bridging generative AI technology and teacher education: Understanding preservice teachers' processes of unit design with ChatGPT. *Contemp. Issues Technol. Teach. Educ.* **2024**, *24*, 1–28.
98. Wyss, L.L.; Hess, R.F. Developing simulation content using ChatGPT to teach nursing students culturally appropriate Amish burn care. *J. Transcult. Nurs.* **2026**, *37*, 169–176. [CrossRef] [PubMed]
99. Yang, F.; Stefaniak, J.E. An exploration of instructional designers' prioritizations for integrating ChatGPT in design practice. *Educ. Technol. Res. Dev.* **2025**, *73*, 2761–2784. [CrossRef]
100. Yao, J.; Liu, M.; Zhang, R.; Zheng, Y. Research on the impact of human-machine collaborative dialogue on normal university students' reflection and instructional design abilities in teaching resources. In *Proceedings of the 2025 5th International Conference on Artificial Intelligence and Education (ICAIE)*; IEEE: New York, NY, USA, 2025; pp. 304–308. [CrossRef]
101. Yu, H.M.; Lee, H.; Kim, S.H. Integrating ChatGPT in teacher education: Examining early childhood and elementary pre-service teachers' AI-assisted math lesson planning. *J. Res. Child. Educ.* **2026**, *40*, 67–89. [CrossRef]
102. Zhang, Y.; Tian, H.; Lu, J. The relationship between higher-order thinking and problem-solving skills development among pre-service teachers using generative AI: An analysis based on moderated mediation. *BMC Psychol.* **2025**, *13*, 1094. [CrossRef] [PubMed]
103. Zhao, Z.J.; Li, J.; Hong, Y.; Yun, T.Y. Lesson Study as an approach to facilitate the integration of Gen-AI into EFL curriculum design. *Int. J. Lesson Learn. Stud.* **2025**, *14*, 253–279. [CrossRef]

104. Pedreschi, D.; Pappalardo, L.; Ferragina, E.; Baeza-Yates, R.; Barabási, A.-L.; Dignum, F.; Dignum, V.; Eliassi-Rad, T.; Giannotti, F.; Kertész, J.; et al. Human-AI coevolution. *Artif. Intell.* **2025**, *339*, 104244. [CrossRef]
105. Holstein, K.; McLaren, B.M.; Aleven, V. Co-designing a real-time classroom orchestration tool to support teacher–AI complementarity. *J. Learn. Anal.* **2019**, *6*, 27–52. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.