Journal Homepage: -www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

Article DOI:10.21474/IJAR01/24578
DOI URL: http://dx.doi.org/10.21474/IJAR01/24578

ISSN (O): 2320-5407
ISSN (P): 3107-4928

RESEARCH ARTICLE

GENERATIVE AI–ENABLED ADAPTIVE LEARNING IN HIGHER EDUCATION: DEEP LEARNING AND DIGITAL COMPETENCE—A CRITICAL NARRATIVE REVIEW

Weam Omran Afifi

1. An-Najah National University, Nablus, Palestine.

Manuscript Info

Manuscript History
Received: 12 September 2026
Final Accepted: 14 October 2026
Published: November 2026

Key words:-
generative artificial intelligence; adaptive learning; deep learning; digital competence; higher education; critical narrative review; scaffolding

Abstract

Generative artificial intelligence (GenAI) offers new possibilities for responsive feedback and individualized support in higher education, but the educational value of a fluent chatbot should not be confused with evidence of durable learning. This critical narrative review examines how GenAI might contribute to deep learning—understood here as conceptual understanding, critical reasoning, and transfer—and digital competence when embedded in genuinely adaptive instructional design s. The review critically interprets published systematic reviews, meta-analyses, conceptual scholarship, and international guidance. It distingu ishes generic conversational assistance from systems that diagnose learner needs, adjust support, progressively withdraw scaffolds, and assess independent performance. The literature indicates benefits for several short-term academic outcomes, alongside unresolved questions about metacognition, self-efficacy, dependence, privacy, and equitable access. We propose a six-phase adaptive scaffolding cycle and an assessment framework aligned with DigComp 2.2. These are design proposals, not empirically validated interventions. We conclude that stronger controlled, longitudinal, and context-sensitive research is needed, including in Palestinian higher education.

"© 2026 by the Author(s). Published by IJAR under CC BY 4.0. Unrestricted use allowed with credit to the author."

Introduction:-

Higher education institutions are confronting a fundamental instructional question: when generative AI can provide a polished answer immediately, how can educators ensure that learners still develop the knowledge and judgment required to produce and evaluate answers independently? The rapid uptake of large language models has created opportunities for explanation, formative feedback, brainstorming, and accessible language support. Yet the same tools can facilitate answer copying, unverified claims, and premature cognitive offloading (Kasneci et al., 2023; Miao & Holmes, 2023).

The central argument of this review is that GenAI becomes educationally consequential through the learning design surrounding it. An unrestricted chat interface is not automatically an adaptive learning environment. Adaptivity requires an interpretable learner model, a mechanism for changing the type or difficulty of support in response to

Corresponding Author:- Weam Omran Afifi
Address:-An-Najah National University, Nablus, Palestine.

evidence, and evaluation that distinguishes assisted performance from independent competence. This distinction is especially important when educational outcomes are described as "deep learning," a term used here in its pedagogical sense rather than as the machine-learning architecture underlying AI models.

The research problem emerges from a persistent mismatch between the visibility of AI-assisted academic products and the invisibility of the learner's independent reasoning. The reviewed syntheses report improvements in performance but provide weaker or inconsistent evidence concerning metacognition and self-efficacy (Deng et al., 2025; Chen & Cheung, 2025). This tension motivates the present inquiry: what would distinguish productive adaptive assistance from a workflow that merely accelerates answer production? The researcher's concern is conceptual and grounded in these published discrepancies, not in unreported field observations or interviews. In particular, there is a need for an instructional account that specifies when support should be offered, how its intensity should change, and what evidence would justify withdrawing it.

The review addresses three questions: (1) What do recent syntheses establish about GenAI and university learning outcomes? (2) What design conditions could connect GenAI-supported adaptivity to deep learning and digital competence? (3) What methodological, ethical, and contextual conditions should guide implementation and future evaluation? Its contribution is a critical synthesis and a testable design framework, not a claim that the proposed intervention has already improved student outcomes.

The distinctive contribution is an explicitly falsifiable alignment among diagnostic evidence, graduated assistance, claim verification, support withdrawal, and unassisted transfer. Rather than presenting six familiar pedagogical principles as a new theory, the review specifies their sequencing and decision points as an integrated design hypothesis. This contribution remains conceptual: its novelty is the proposed combination and operational specification, not proof of superior effectiveness.

Research problem, rationale, and contribution:-

The research problem is not whether generative AI can produce useful educational responses, but whether performance achieved with such responses represents learning that students can reproduce, justify, and transfer without assistance. The author identifies a conceptual and evaluative mismatch in the reviewed literature: studies often report task achievement or perceptions of usefulness, while the instructional mechanisms responsible for independent conceptual understanding, metacognitive monitoring, and digital judgment are less clearly isolated. This concern arises from critical comparison of published evidence, not from unreported observations or interviews at the author's university.

The specific gap is the absence, within the sources reviewed here, of a sufficiently explicit bridge between (a) diagnosis-contingent GenAI assistance, (b) a decision rule for fading that assistance, (c) verification of AI claims as a digital-competence practice, and (d) delayed unassisted transfer as the criterion for success. The review therefore asks how these components can be organized into a coherent, falsifiable instructional design rather than treating access to a chatbot as adaptivity. The contribution is a proposed instructional and assessment architecture—not a newly discovered learning principle or a validated technological system. Its distinctive claim is that a GenAI intervention should count as educationally adaptive only when support decisions are linked to observable learner evidence and independently assessed outcomes. The proposed architecture is intended to generate testable comparisons with unguided GenAI use and conventional teaching.

Review approach and boundaries:-

This article is a critical narrative review, not a systematic review or meta-analysis. Its purpose is to integrate conceptually relevant, identifiable scholarly syntheses and policy frameworks, to expose inferential gaps, and to develop a pedagogical model. Sources were selected purposively for relevance to higher education, GenAI learning outcomes, chatbot use, AI implementation, digital competence, and governance. Priority was given to peer-reviewed syntheses with traceable publisher records and to primary institutional frameworks. Publisher and institutional records were consulted to verify bibliographic identity. No exhaustive database search, preregistered protocol, duplicate screening, quality scoring, or pooled effect estimation was conducted; therefore no claim of comprehensiveness or systematic selection is made.

The reviewed sources are heterogeneous. Some examine artificial intelligence broadly rather than generative AI; others assess ChatGPT rather than adaptive systems; and policy frameworks prescribe competencies or safeguards rather than test learning outcomes. Their conclusions are therefore compared by evidential scope instead of being

treated as interchangeable causal proof. The framework developed below is an interpretive proposal derived from this synthesis.

Analytic procedure:-
Each source was read for (a) population and technology scope, (b) outcome type, (c) evidence of adaptivity, (d) evidence of independent or delayed assessment, and (e) implications for digital competence and safeguards. Cross-source comparisons were used to distinguish observed outcomes from plausible mechanisms. This is an interpretive analytic structure rather than a reproducible systematic coding protocol; no formal inter-rater reliability or risk-of-bias assessment was undertaken.

Conceptual foundations:-
Deep learning as an educational outcome:-
Deep learning involves explaining relationships, integrating evidence, interrogating assumptions, and transferring knowledge to unfamiliar problems. It cannot be inferred from the grammatical quality of AI-assisted essays alone. A student may submit a sophisticated product while remaining unable to defend its reasoning, reproduce its logic, or apply the concept without assistance. Appropriate indicators include concept explanations, justified decisions, transfer tasks, delayed retention, and reflective error correction. Such measures should be gathered independently of the AI system when the purpose is to estimate learner competence.

What makes an environment adaptive:-
A conversational assistant responds to a prompt; an adaptive instructional environment responds to evidence about the learner. An operational definition requires at least four elements: diagnosis of current understanding; conditional selection of explanations, hints, or challenge level; repeated formative checks; and explicit fading of support as mastery grows. A fifth element—human oversight—is essential when learner profiling or consequential decisions are involved. Without these components, describing an intervention as "adaptive" risks overclaiming its pedagogical sophistication.

Digital competence beyond tool operation:-
DigComp 2.2 defines five areas of digital competence: information and data literacy; communication and collaboration; digital content creation; safety; and problem solving (Vuorikari et al., 2022). For GenAI-supported education, these domains imply more than the ability to formulate prompts. Students need to verify claims, understand limitations of synthetic content, attribute assistance appropriately, protect sensitive data, evaluate bias, and choose when not to use automation. An intervention may improve productivity without producing these transferable competencies.

Evidence from prior research:-
Broad AI and chatbot evidence:-
Crompton and Burke (2023) reviewed 138 articles on artificial intelligence in higher education published across 2016–2022. Their mapping is valuable for situating tutoring, assessment, and institutional uses of AI, but the period largely predates the widespread availability of contemporary GenAI chatbots. Its findings therefore provide historical context rather than direct estimates of GenAI effects. Labadze et al. (2023) systematically reviewed educational AI chatbots, identifying opportunities for responsive support alongside implementation challenges. The broad chatbot category includes technologies with different capabilities, so results should not be assumed to generalize to generative adaptive tutoring.

Learning outcomes and the limits of aggregate effects:-
Deng et al. (2025) synthesized 69 experimental studies of ChatGPT in education and reported positive effects on academic performance, affective-motivational states, and higher-order thinking propensities, while finding no improvement in self-efficacy. The authors also reported reduced mental effort. Reduced effort may reflect useful efficiency or the outsourcing of productive cognitive work; it cannot by itself establish deep learning. Crucially, a meta-analysis of ChatGPT interventions is not a meta-analysis of continuously adapting learner-model systems.
Chen and Cheung (2025) examined generative AI and university learning outcomes in a systematic review and meta-analysis. Their findings support benefits for several outcomes, including academic achievement and higher-order thinking, but no statistically significant effect on metacognition. This contrast challenges the assumption that stronger task performance necessarily entails better self-monitoring. Because interventions, outcomes, and comparison conditions vary, aggregate benefits should guide hypotheses rather than serve as universal predictions for any particular campus or course.

Implementation in authentic university settings:-

Belkina et al. (2025) synthesized case studies of GenAI implementation across higher education. Their review is especially useful because it examines how applications are embedded in educational activities and classifies them through pedagogical frameworks. However, implementation case studies do not uniformly establish causal effects or long-term retention. The distinction between feasibility, student satisfaction, and demonstrable learning gains must remain explicit. Together with the outcome meta-analyses, this literature suggests a design problem: GenAI can assist learning activities, but the evidence is thinner on whether particular adaptive mechanisms preserve autonomy and transfer.

Cross-study interpretation:-

Crompton and Burke (2023) establish the breadth of earlier AI applications but cannot resolve the effectiveness of present-day GenAI. Labadze et al. (2023) address chatbots across architectures, leaving the adaptive mechanism unspecified. Deng et al. (2025) and Chen and Cheung (2025) provide stronger aggregate outcome evidence, yet their findings on self-efficacy and metacognition complicate a simple performance-to-learning inference. Belkina et al. (2025) offer implementation insight without consistently isolating causal mechanisms. Accordingly, these studies complement rather than replicate one another: the reviews establish opportunity and uncertainty, while the present article targets the missing linkage between specific instructional decisions and independently demonstrated competence. This is a design-oriented contribution, not an additional effect-size estimate.

A comparative evidence matrix:-

Source and evidence	Supported contribution	Boundary for this review
Crompton & Burke (2023): systematic review; 138 articles	Maps AI use in higher education before widespread GenAI adoption	Not direct evidence for contemporary GenAI adaptivity
Labadze et al. (2023): chatbot review	Synthesizes educational chatbot uses and challenges	Chatbots are not uniformly generative or adaptive
Deng et al. (2025): meta-analysis; 69 experiments	Positive effects on several learning outcomes; no self-efficacy improvement	Most interventions do not isolate adaptive scaffolding
Chen & Cheung (2025): meta-analysis	Benefits across several outcomes; metacognition not significantly improved	Aggregated outcomes do not establish durable transfer
Belkina et al. (2025): case-study review	Describes practical implementation across settings	Case studies alone cannot establish causality
Vuorikari et al. (2022): competence framework	Five digital competence domains	Normative framework, not an intervention trial
Miao & Holmes (2023): UNESCO guidance	Human-centred governance, privacy and validation	Policy guidance, not evidence of effect size

Critical comparison and the position of the present review

The reviewed publications answer different questions and cannot be arranged as if they were successive demonstrations of one intervention. Crompton and Burke (2023) map the historical AI-in-higher-education field, whereas Labadze et al. (2023) synthesize heterogeneous chatbot applications; neither identifies the effect of a specified GenAI hint-fading rule. Deng et al. (2025) and Chen and Cheung (2025) provide stronger aggregated outcome evidence, but their intervention categories do not establish that improvements persist when AI is removed. Belkina et al. (2025) contributes implementation insight without uniformly providing causal or delayed-outcome evidence. DigComp 2.2 and UNESCO guidance specify desired capabilities and safeguards rather than empirically testing the model proposed here.

These contrasts reveal two competing interpretations of positive performance findings. The learning interpretation is that well-timed explanations help students reorganize concepts and later solve unfamiliar tasks. The substitution interpretation is that students delegate reasoning to a fluent system and improve only the assisted artifact. The available syntheses do not reliably adjudicate between these interpretations for the specific adaptive mechanism proposed in this review. This is the precise inferential boundary the present article seeks to make visible. Unlike the cited outcome reviews, the present article does not estimate a mean effect. Unlike a general implementation

framework, it specifies a sequence of learner evidence, conditional help, verification, and independent reassessment. Unlike a policy framework, it proposes measurable instructional decisions. These differences define its conceptual contribution while also limiting its empirical claims.

Critical synthesis: mechanisms and competing explanations:-

Feedback can support reasoning or replace it:-

An adaptive tutor could use diagnostic errors to select a question that exposes a misconception, then offer progressively more explicit hints. This is a plausible route to conceptual restructuring. Yet a system that simply supplies the final answer after the first difficulty may improve assignment completion while bypassing reasoning. The instructional unit of analysis should therefore be the sequence of learner actions, prompts, feedback, revision, and subsequent independent response—not merely the quality of the final artifact.

The metacognition problem:-

The non-significant metacognitive findings reported by Chen and Cheung (2025), considered alongside the self-efficacy result in Deng et al. (2025), caution against assuming that more successful outputs imply greater self-regulation. Metacognitive growth requires explicit monitoring tasks: confidence judgments before feedback, error explanations, comparison of alternative answers, and retrospective analysis of when AI advice was misleading. These tasks should be scored independently rather than inferred from student perceptions of convenience.

Digital competence requires critical use:-

Prompting skill is only one narrow element of digital competence. Students should be able to check an AI-generated citation against a publisher record, identify missing context, explain why a claim is uncertain, and avoid sharing personal or institutional data with inappropriate services. Digital competence can be assessed with scenario-based tasks mapped to the five DigComp domains. A productive intervention might deliberately present a plausible but unsupported AI claim and evaluate the student's verification process; such testing must be transparent and ethically designed.

Equity and local transferability:-

Benefits demonstrated in one language, discipline, or institutional environment should not be presumed to transfer automatically to Palestinian universities. Access to reliable connectivity, suitable devices, accessible interfaces, Arabic-language performance, faculty preparation, and institutional data policies are contextual variables to investigate—not facts that this review has measured at An-Najah National University. Research in the Palestinian context should begin with a local needs assessment and stakeholder consultation rather than a universal technology prescription.

The proposed mechanism is consistent with sociocultural accounts of learning in which assistance is contingent on what a learner can accomplish with support, and with scaffolding perspectives that emphasize eventual withdrawal (Vygotsky, 1978; Wood et al., 1976). Self-regulated learning theory further suggests that planning, monitoring, and reflection must remain learner activities rather than be delegated to a chatbot (Zimmerman, 2002). These theories support design plausibility, not efficacy. At least three competing explanations should be tested: improvements may arise from extra time on task, more fluent output without conceptual change, or increased access to feedback rather than from adaptive fading itself. Independent transfer, delayed assessment, and logs of hint use are necessary to discriminate among these accounts.

Context-sensitive implementation in Palestinian higher education:-

For Palestinian universities, the framework should be treated as a context-sensitive design option rather than a ready-to-deploy intervention. Before adoption, an institution would need to examine students' access to devices and connectivity, disability-related accessibility requirements, the availability of approved AI services, and the adequacy of Arabic and English instructional responses. These are proposed assessment domains, not findings about the conditions at An-Najah National University or other Palestinian institutions.

A feasible pilot could begin in a single voluntary course with an instructor-selected concept, a small bank of rubric-scored diagnostic and transfer tasks, and a bilingual set of vetted prompts. Students should have an equivalent non-AI learning route. Instructors and students could jointly review whether the AI explanations use accurate terminology in both languages and whether source-verification tasks can be completed with accessible library resources. Institutional approval and applicable privacy safeguards would precede any collection of learner interaction data. The local contribution of such a pilot would be to identify which assumptions in the proposed cycle

hold across language, discipline, and access conditions. Any claims about Palestinian student outcomes would require locally collected evidence, appropriate consent and oversight, and reporting of limitations; none is supplied by the present review.

Proposed six-phase adaptive scaffolding cycle:-

The following cycle is an original instructional design proposal synthesized from the reviewed literature and guidance. It has not been empirically validated in this article. Mechanism and design logic. The cycle is organized as a feedback-controlled sequence: an unassisted baseline establishes an observable starting point; a learner-generated attempt makes reasoning visible; hints respond to errors rather than merely to requests; verification tests epistemic judgment; an unassisted transfer task tests whether support has become internalized; and human review determines whether to fade, repeat, or intensify support. Advancement should be conditional on predeclared rubric thresholds rather than on conversational fluency. The model is not derived from newly collected results: it is a hypothesis informed by the pattern and limitations of published findings, especially the gap between performance gains and self-regulatory outcomes.

Derivation, distinctive mechanism, and decision logic:-

The six phases are derived as a reasoned design response to tensions identified in the evidence, not statistically extracted from the reviewed studies. Reports of improved performance but uncertain metacognition motivate unassisted diagnosis and reflective calibration (Deng et al., 2025; Chen & Cheung, 2025). The heterogeneity of chatbot applications motivates conditional rather than generic assistance (Labadze et al., 2023). DigComp 2.2 motivates explicit verification and responsible use (Vuorikari et al., 2022); human-centred safeguards motivate instructor review (Miao & Holmes, 2023). None of these sources directly validates the six-phase sequence.

The proposed innovation lies in coupling three decision gates: an entry gate based on unassisted evidence of a misconception; an assistance gate that permits a more explicit hint only after a learner attempt or diagnostic response; and an exit gate requiring an unassisted, novel task before support is considered successfully faded. If the exit task reveals a persistent misconception, the instructor returns the learner to targeted practice rather than recording assisted completion as mastery. These are proposed rules for future operationalization, not properties of an existing deployed system. The model therefore differs from a linear checklist: the sequence contains conditional loops, an explicit separation between supported and unsupported performance, and a second outcome stream for critical digital competence. Its added value would be falsified if learners receiving this structure showed no advantage over unguided GenAI access on delayed, unassisted transfer or verification tasks.

Phase 1 — Unassisted diagnosis:-

Students complete a baseline concept and transfer task without AI. Teachers identify misconceptions and set explicit learning goals.

Phase 2 — Task framing and inquiry:-

Learners state a question, assumptions, evidence needs, and a preliminary answer before requesting AI help.

Phase 3 — Graduated adaptive hints:-

The system provides a conceptual cue, then a targeted sub-question, and only later a partial worked example if diagnostic evidence warrants it. Operational decision rule (proposed): the first response should elicit a learner explanation; a conceptual cue follows an incorrect or incomplete explanation; a targeted question follows a persistent misconception; and a partial worked example is reserved for repeated difficulty. A new independent check then determines whether to reduce assistance. Thresholds, such as the number of errors triggering a hint, must be calibrated and preregistered in future trials; no validated cutoff is claimed here.

Phase 4 — Evidence verification:-

Students check AI claims against credible sources, document corrections, and explain uncertainty or bias.

Phase 5 — Independent synthesis and transfer:-

Students complete a new, comparable task without AI and justify their reasoning in writing or orally.

Phase 6 — Reflection and human review:-

Students compare baseline and independent performance, evaluate reliance on hints, and receive instructor feedback; assistance is faded or recalibrated.

Distinctiveness and testable propositions:-

The distinctive feature is the coupling of two gates: a learning gate (unassisted explanation and transfer) and a digital-competence gate (source checking, uncertainty identification, and safe handling of information). A student should not be classified as independently competent solely because an AI-assisted product is correct.

Proposition 1: compared with unguided GenAI, diagnostic hints with fading will produce stronger delayed unassisted transfer.

Proposition 2: explicit verification tasks will improve performance-based digital competence more than prompting practice alone.

Proposition 3: any immediate performance advantage may shrink when AI access is removed. All three propositions remain untested.

Evidence-to-design traceability:-

Phase 1 (unassisted diagnosis): addresses the risk that polished assisted work conceals misconceptions; proposed indicator: baseline explanation and transfer rubric. Phase 2 (inquiry framing): addresses passive answer consumption; indicator: learner-generated assumptions and provisional reasoning. Phase 3 (conditional hints): addresses undifferentiated chatbot help; indicator: attempt–hint–revision sequence and level of help required. Phase 4 (verification): operationalizes information literacy and safety; indicator: independently checked claims, source validity, and attribution. Phase 5 (unassisted transfer): discriminates learning from substitution; indicator: novel-task score without AI and delayed retention. Phase 6 (reflection and review): addresses uncertainty about self-monitoring; indicator: confidence–accuracy calibration, justified revisions, and documented instructor decisions.

The links above are theoretical design warrants inferred from the literature, not findings that each phase has already caused improvement. For empirical testing, the six-phase package should be compared with a matched-content condition offering unguided GenAI, with exposure time and instructor support recorded.

Illustrative classroom workflow and technical feasibility:-

Consider a hypothetical university research-methods lesson on distinguishing correlation from causation. First, students explain an unfamiliar claim without AI; an instructor scores the explanation against a short misconception rubric (Phase 1). Students then formulate a tentative interpretation and list the evidence they would need (Phase 2). An approved GenAI interface is constrained to ask for the student's reasoning before offering a conceptual cue, then a targeted question, and only after further difficulty a partial worked example (Phase 3). Students verify the AI's methodological claims against course readings and document any correction (Phase 4). They next evaluate a different case without AI (Phase 5), and the instructor compares baseline and transfer responses, reviews hint use, and decides whether further support is needed (Phase 6). This is an illustrative design scenario, not an implemented lesson or observed result.

Technically, the minimum viable implementation does not require a newly trained language model. It could combine a university-approved AI interface, instructor-authored task and hint templates, a simple rubric-based record of misconception categories, and a human-reviewed decision log. A more automated version might map rubric outcomes to hint levels, but its classifications would need validation against instructor judgments before being used for consequential decisions. No validated accuracy, cost, scalability, or system reliability is claimed.

Feasibility should be assessed before any pilot through four checks: (1) pedagogical fidelity—whether the tool withholds full solutions and fades assistance as intended; (2) technical reliability—whether prompts, logs, accessibility features, and bilingual responses function consistently; (3) governance—whether data minimization, retention rules, consent, and human override are in place; and (4) resource burden—whether instructors can maintain rubrics, verify AI outputs, and support students without excessive workload. Where these conditions cannot be met, instructor-led scaffolding is a preferable comparator or alternative. For a future evaluation, implementation fidelity could be reported as the proportion of interactions that follow the prescribed attempt-before-hint sequence, the distribution of hint levels, and the proportion of exit assessments completed without AI.

These are proposed process indicators, not measured outcomes. The central empirical test remains whether the structured approach improves delayed independent transfer and critical verification relative to equally resourced unguided GenAI access.

Assessment and evaluation design:-

A credible evaluation should separate three outcomes: immediate task performance, durable deep learning, and digital competence. Immediate performance can be assessed with rubric-scored products, but durable learning requires delayed or novel transfer tasks completed without AI. Digital competence requires performance-based

evidence across DigComp domains, such as source evaluation, secure information handling, attribution, and troubleshooting. Self-report questionnaires may supplement, but should not replace, demonstrated performance.
A future quasi-experimental study could compare (A) GenAI with diagnostic scaffolding and fading, (B) unguided GenAI access, and (C) conventional instruction where feasible. Baseline equivalence, instructor effects, prior AI literacy, disciplinary differences, access conditions, and contamination between groups would need explicit management. Preregistered outcomes, blinded rubric scoring, transparent missing-data treatment, and delayed post-tests would strengthen inference. This design is proposed for future research; no participants were recruited and no data were collected for the present article.

A theory-linked measurement strategy should capture conceptual change through explanation rubrics and unfamiliar transfer problems; self-regulation through pre-feedback confidence estimates and error-monitoring tasks; and digital competence through scenario-based judgments mapped to the five DigComp 2.2 areas (Vuorikari et al., 2022). The study should also measure prompt frequency, hint level, and the difference between assisted and unassisted scores. A preregistered mediation analysis could explore whether reduced reliance on hints accompanies transfer gains, but such an analysis would not establish mechanism without adequate design and assumptions.

Testable propositions and interpretation of possible results:-
Proposition 1: compared with unguided GenAI access, diagnosis-contingent scaffolding with fading will improve delayed unassisted transfer after controlling for baseline competence. Proposition 2: explicit verification tasks will improve performance on independently scored AI-claim evaluation scenarios. Proposition 3: assisted assignment quality will correlate imperfectly with unassisted transfer, revealing cases of performance without durable learning. These are hypotheses for future studies, not results of this review. A positive immediate score with a null delayed-transfer effect would support a productivity or substitution explanation rather than a durable-learning claim. Better verification without improved subject-matter transfer would indicate that digital competence and disciplinary understanding can develop separately. Null or negative effects would require examination of hint timing, prior knowledge, language quality, access, and instructor implementation rather than post hoc claims that the model necessarily works.

Ethical and institutional safeguards:-
UNESCO guidance emphasizes a human-centred approach, institutional validation, and data protection (Miao & Holmes, 2023). In practice, a university should specify permissible AI use, protect learner records, avoid entering confidential material into unapproved systems, provide meaningful human review of consequential judgments, and ensure non-AI alternatives where access is unequal. Staff development should address pedagogical prompting, verification, assessment redesign, and accessibility. An institution should also periodically test model performance across Arabic and English tasks rather than assume linguistic equivalence.

Comparison with a published IJAR conceptual-framework article:-
For journal-level positioning, Chaoui and Derkaoui (2025) offer a conceptual framework integrating design thinking and ethical AI for Moroccan teacher-researcher development. Their article foregrounds a country-specific training problem and contextual analysis. The present review addresses a different population and outcome: students' independent learning and digital competence under GenAI-supported instruction. The relevant lesson is methodological rather than a claim of superiority: a conceptual framework should specify its context, components, derivation, and pathway to empirical validation. The present model accordingly adds explicit diagnostic and exit criteria and limits its claims to a researchable proposal.

Limitations and future Research:-
This narrative review is selective and interpretive. It did not systematically identify all relevant studies, appraise their risk of bias, or pool effects. Its cited empirical syntheses investigate different populations, technologies, interventions, and outcome measures. The proposed six-phase model is a conceptual contribution, not evidence of effectiveness. Moreover, the review does not present empirical findings from Palestine or An-Najah National University. Future studies should isolate the effect of adaptive scaffolding from the effect of access to GenAI; measure unassisted transfer and delayed retention; evaluate metacognitive calibration and student agency; examine accessibility and Arabic-language performance; and report privacy, governance, and cost implications. Mixed-method and longitudinal designs could explain not only whether an intervention works but for whom, under which constraints, and through which learning processes.

Future research should also report null and adverse outcomes, including overreliance, erroneous AI guidance, differential access, and faculty workload. Comparisons must distinguish a merely responsive chatbot from an adaptive system with documented learner-state rules. The conceptual framework may require revision if independent learning gains do not follow from fading, or if verification tasks create burdens without measurable competence gains.

Conclusion:-

The available literature supports cautious optimism about GenAI as an educational resource, but not a blanket claim that chatbots produce deep learning or digital competence. The decisive issue is instructional design: diagnostic assessment, calibrated feedback, verification, independent application, and accountable human oversight. The six-phase cycle proposed here translates those principles into a researchable model. Its value must be tested rather than presumed, particularly in context-sensitive higher education environments. The article therefore contributes an operational research agenda rather than an empirically established model. Its central claim is conditional: GenAI-supported adaptivity merits educational endorsement only when learners can demonstrate what they understand, verify, and transfer after assistance is removed.

References:-

1. Belkina, M., Daniel, S., Nikolic, S., Haque, R., Lyden, S., Neal, P., Grundy, S., & Hassan, G. M. (2025). Implementing generative AI (GenAI) in higher education: A systematic review of case studies. Computers and Education: Artificial Intelligence, 8, 100407. https://doi.org/10.1016/j.caeai.2025.100407
2. Chaoui, S., & Derkaoui, K. (2025). Towards innovative continuous training of teachers-researchers in Morocco: A conceptual framework integrating design thinking and ethical artificial intelligence. International Journal of Advanced Research, 13(8), 161–186. https://doi.org/10.21474/IJAR01/21499
3. Chen, S., & Cheung, A. C. K. (2025). Effect of generative artificial intelligence on university students learning outcomes: A systematic review and meta-analysis. Educational Research Review, 49, 100737. https://doi.org/10.1016/j.edurev.2025.100737
4. Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. International Journal of Educational Technology in Higher Education, 20, Article 22. https://doi.org/10.1186/s41239-023-00392-8
5. Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. Computers & Education, 227, 105224. https://doi.org/10.1016/j.compedu.2024.105224
6. Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274
7. Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: Systematic literature review. International Journal of Educational Technology in Higher Education, 20, Article 56. https://doi.org/10.1186/s41239-023-00426-1
8. Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO. https://doi.org/10.54675/EWZM9535
9. Vuorikari, R., Kluzer, S., & Punie, Y. (2022). DigComp 2.2: The Digital Competence Framework for Citizens—With new examples of knowledge, skills and attitudes. Publications Office of the European Union. https://doi.org/10.2760/115376
10. Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.
11. Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. Journal of Child Psychology and Psychiatry, 17(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x
12. Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. Theory Into Practice, 41(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2