Journal of Informatics Education and Research
ISSN: 1526-4726
Vol 6 Issue 1 (2026)

# Artificial Intelligence as a Cognitive Scaffold in First Year Computing Education

**Eric Howard[1,2], Hardique Dasore[1], Shah Haque[1], Raddhika Kuttala[1], Mohamad Mahmoud Al Zein[1], Hom Nath Dhungana[1]**

[1]Southern Cross Institute
[2]Macquarie University

**Abstract**

The rapid adoption of artificial intelligence tools in higher education has raised significant questions about their role in student learning, particularly in introductory computing courses. While concerns persist that AI may replace conceptual understanding or encourage surface learning, this paper argues that artificial intelligence can function effectively as a cognitive scaffold when appropriately integrated into first year computing education. Drawing on cognitive load theory, scaffolding theory, and contemporary educational technology research, this paper proposes a framework for ethical and pedagogically aligned AI use that supports novice learners without undermining foundational conceptual development. The paper analyses the benefits, risks, and design principles required to ensure AI acts as a support mechanism rather than a substitute for learning.

## 1 Introduction

First year computing education presents a well documented challenge for both students and educators. Novice learners are expected to develop abstract thinking skills while also acquiring a specialised technical vocabulary and internalising problem solving strategies that are unfamiliar, cognitively demanding, and often counterintuitive to prior learning experiences. In many introductory units, students must learn to translate ambiguous problem statements into precise computational steps, reason about control flow and state changes, and connect symbolic representations to executable behaviour. These demands arrive at the same time that learners are still building basic fluency with the programming environment, the mechanics of syntax, and the social practices of computing such as debugging and iterative refinement. As a result, early learning can become dominated by confusion and fragmented trial and error rather than coherent conceptual growth. High attrition rates in introductory computing units are frequently attributed to cognitive overload, reduced confidence, and persistent difficulty bridging the gap between conceptual explanations and practical implementation within authentic tasks.

The emergence of generative artificial intelligence tools introduces new possibilities for addressing these challenges. When used appropriately, AI systems can provide immediate feedback, offer alternative explanations tailored to novice comprehension, and guide problem decomposition by helping learners break complex tasks into smaller and more manageable steps. In principle, this can reduce unnecessary friction during early learning, especially when students encounter barriers such as unclear instructions, limited feedback cycles, or difficulty articulating their misunderstandings. At the same time, unregulated or poorly designed use of AI risks replacing student reasoning with automated output, encouraging acceptance of generated solutions without understanding, and weakening the development of foundational skills that introductory computing is intended to build. This paper positions artificial intelligence as a cognitive scaffold that supports learning during the earliest stages of skill acquisition while preserving the primacy of conceptual understanding. The central argument is that AI can function as a temporary support mechanism that reduces extraneous cognitive burden and sustains productive progress, provided that its use is explicitly structured to promote explanation, reflection, and independent synthesis rather than task completion.

A further motivation for this work is the shift in the learning ecology of introductory computing. Students now encounter AI enabled assistance not only within formal study settings but also within informal learning channels, peer communities, and commercial platforms that foreground rapid solution generation. This changes the baseline assumptions that underpinned many established teaching and assessment designs. In particular, the availability of fluent generated code, explanations, and debugging suggestions can reduce the natural scarcity of feedback that historically shaped novice learning. At the same time, ubiquitous assistance can compress the space for productive struggle that is often necessary for developing robust mental models of execution, control flow, state, and abstraction. The pedagogical question is therefore not limited to whether AI use should be permitted, but rather how to design learning environments where AI use supports the formation of conceptual structures that remain stable when external support is reduced.

This paper adopts the premise that first year computing should be defined by the development of durable thinking skills rather than by short term artefact production. In introductory programming, student success depends on learning to reason about the behaviour of programs, including how variables evolve over time, how conditional branches partition execution paths, how loops accumulate effects, and how functions encapsulate and compose behaviour. These skills require learners to coordinate multiple interacting elements under working memory constraints. If instructional design fails to manage these constraints, learners may default to surface level pattern replication, trial and error debugging, or reliance on external authorities. AI can either amplify these failure modes by supplying polished outputs that obscure reasoning, or it can mitigate them by providing targeted explanatory prompts that reduce extraneous burden while preserving germane cognitive effort invested in schema construction.

A key contribution of the paper is to formalise a boundary between AI as a scaffold and AI as a substitute. In educational terms, a scaffold is a temporary structure that enables performance that is currently beyond independent capability, contingent on learner needs, and deliberately withdrawn as competence increases. The risk with generative AI is that the same system can simultaneously provide supportive prompts and fully formed solutions, often without clear signalling about which mode is being invoked. This ambiguity is consequential in first year computing because many students do not yet possess the epistemic skills required to evaluate whether an explanation is correct, whether a solution generalises, or whether a suggested approach preserves the intended constraints of a task. The framework proposed here treats AI output as provisional and subject to verification, positioning evaluation, testing, tracing, and justification as core learning practices rather than optional add ons.

The argument is also anchored in an integrity sensitive view of learning design. Concerns about misconduct, over assistance, and unfair advantage are often framed as behavioural problems, but they are also symptoms of misalignment between assessment constructs and the tools available to students. If an assessment primarily measures whether a student can produce a plausible artefact outside a controlled setting, then high quality generated output can undermine the validity of the measure. A scaffold oriented integration reframes this problem by shifting emphasis toward reasoning evidence, such as explanations of behaviour, justification of design choices, analysis of trade offs, and demonstration of verification practice. Under this framing, AI use is not automatically incompatible with academic standards, but it must be integrated into tasks that elicit evidence of cognition rather than evidence of completion.

This paper therefore addresses three guiding questions that arise directly from contemporary first year computing practice. First, what theoretical lenses best explain why novices struggle and where AI can reduce extraneous burden without reducing germane learning effort. Second, what design principles and boundaries are required so that AI use promotes explanation, reflection, and independent synthesis rather than substitution. Third, what risks are most likely to emerge in real classrooms, and what mitigation strategies can be embedded in curriculum, assessment, and student guidance to preserve learning outcomes and equity. The remainder of the paper develops these questions through cognitive load theory and scaffolding theory, then proposes practical principles for ethical AI scaffolding, and finally outlines risks, mitigation strategies, and research directions for evaluating effectiveness in diverse cohorts.

The scope of the paper is introductory computing education, with particular relevance to first year programming units where learners are acquiring both conceptual foundations and operational fluency. The focus is not on developing new AI algorithms, but on articulating a pedagogically aligned integration model that treats AI as a cognitive scaffold embedded in learning design. The intended outcome is a framework that supports novice learners in developing the conceptual and metacognitive capabilities that computing education requires, while maintaining clear boundaries that protect the integrity and validity of learning and assessment.

## 2 Theoretical

Cognitive load theory provides a useful lens for analysing why introductory computing can be difficult for novice learners and how instructional interventions may improve learning outcomes. The theory distinguishes between intrinsic cognitive load, extraneous cognitive load, and germane cognitive load. Intrinsic load is associated with the inherent complexity of the material and the number of interacting elements that must be processed simultaneously. Extraneous load is imposed by the manner in which information is presented, including poor task design, unclear instructions, or confusing interfaces. Germane load refers to the cognitive effort invested in constructing and refining mental models, including schema formation and transfer of understanding to new contexts.

First year computing students frequently experience high intrinsic load due to the abstract nature of core concepts such as algorithms, data structures, procedural and object oriented reasoning, control flow, and system level logic. Many of these ideas require learners to reason about invisible states, symbolic representations, and layered levels of abstraction while still developing basic fluency with syntax and development tools. The intrinsic demands increase further when tasks require integration of multiple concepts, such as combining conditional logic with iteration, function composition, and data structure manipulation. In this environment, poorly designed instruction can substantially increase extraneous load. Examples include assessments that assume unstated prior knowledge, laboratories that introduce multiple new tools simultaneously, or learning resources that provide incomplete explanations that force novices into unproductive trial and error. When extraneous load is high, cognitive capacity is diverted away from meaningful learning and toward coping with confusion, resulting in fragmented knowledge and reduced persistence.

Artificial intelligence tools can be conceptualised as mechanisms for reducing extraneous load when they are used to support comprehension rather than replace thinking. In a first year context, AI can clarify task instructions, reframe definitions using novice appropriate language, provide worked examples that highlight reasoning steps, and support incremental exploration of a problem by encouraging learners to focus on one conceptual component at a time. AI can also assist learners in identifying where a misunderstanding occurred by prompting them to articulate their reasoning and by highlighting conceptual gaps. When such support is aligned with learning objectives, it can preserve cognitive resources for germane processing, enabling students to invest effort in building robust conceptual schemas rather than relying on rote execution or surface level pattern matching.

Scaffolding theory describes how learners can be supported through temporary instructional structures that enable them to perform tasks that would otherwise be beyond their current level of competence. Scaffolds are intended to be contingent on learner needs, calibrated to the learners zone of proximal development, and gradually withdrawn as competence increases. In computing education, scaffolding is frequently operationalised through guided laboratory activities, partially completed templates, stepwise instructions, algorithmic outlines, formative feedback cycles, peer support structures, and instructor prompted reflection. These supports can help novices manage complexity, sustain motivation, and develop correct mental models while they transition toward independent problem solving.

Artificial intelligence extends traditional scaffolding mechanisms by offering adaptive and on demand support that can respond immediately to learner queries. Unlike static resources, AI systems can rephrase explanations, provide multiple representational forms of the same concept, and adjust the granularity of guidance based on what a learner asks. For example, a learner may request a simpler explanation, a conceptual analogy, or a stepwise reasoning trace that connects problem requirements to computational operations. AI can also support metacognitive scaffolding by encouraging learners to explain their approach, justify choices, and reflect on why an outcome occurred. In addition, AI can function as a procedural scaffold by assisting learners in planning, decomposing tasks, and identifying the next actionable step when they are stuck.

The critical requirement is that scaffolds must not complete cognitive work on behalf of the learner. If AI provides full solutions without requiring reasoning, the scaffold becomes substitutive rather than supportive, and the student may progress through tasks without developing durable understanding. Effective scaffolding therefore requires careful boundary setting, including prompt design that prioritises explanation over answer generation, learning activities that require students to articulate reasoning, and assessment designs that evaluate conceptual understanding rather than the ability to reproduce outputs. Equally important is scaffold withdrawal, where the level of AI support is intentionally reduced over time as learners develop confidence and independent competence. In this way, AI can be positioned as a temporary cognitive partner that strengthens foundational learning rather than a mechanism that bypasses it.

When used as a cognitive scaffold, artificial intelligence can support novice learners in a range of targeted and pedagogically valuable ways that are particularly relevant to first year computing education. At this stage of study, students are often capable of engaging with concepts but struggle to organise their thinking, interpret problem requirements, or connect abstract ideas to concrete actions. AI can assist by acting as an intermediary layer between instructional material and learner cognition, helping students navigate complexity without removing the need for reasoning.

One key support function is the explanation of abstract computing concepts in alternative forms. Many foundational ideas in computing are inherently abstract and symbolic, requiring learners to reason about processes that are not directly observable. AI systems can rephrase explanations using simpler language, analogies, or stepwise reasoning, allowing students to approach the same concept from multiple perspectives. This multiplicity of representations can help learners form more robust mental models, particularly when initial explanations fail to resonate.

Another important function is guided problem decomposition. Novice learners frequently struggle to break complex tasks into manageable components, leading to cognitive overload and unproductive trial and error. AI can prompt learners to identify subproblems, clarify assumptions, and sequence steps in a logical order. When designed appropriately, this guidance supports planning and organisation without executing the task itself. This form of scaffolding is especially valuable in early computing tasks where students must learn how to think computationally before they can implement solutions effectively.

AI can also provide feedback that focuses on reasoning processes rather than final answers.

Traditional feedback mechanisms in large classes are often delayed or limited in scope, which can leave students uncertain about whether their approach is conceptually sound. AI based feedback can encourage learners to articulate their reasoning, reflect on decision making, and identify misconceptions. By responding to how a learner thinks rather than what they produce, AI reinforces the importance of understanding over correctness alone.

A further support function lies in helping students debug misconceptions rather than code in isolation. Many novice errors arise not from syntax issues but from incorrect mental models about how systems behave. AI can assist learners in examining why an outcome occurred, what assumptions led to a particular result, and how alternative reasoning might change the behaviour of a system. This shifts the focus from surface level correction to deeper conceptual repair, which is essential for long term learning.

For example, AI can assist students in understanding why an algorithm behaves in a particular way by prompting them to trace logic step by step, consider edge cases, or explain state changes over time. Crucially, this can be achieved without providing a complete solution. By withholding final answers and instead guiding reflection, AI preserves the learners responsibility for synthesis and decision making while reducing frustration, uncertainty, and disengagement that often characterise early computing experiences.

A central challenge in integrating artificial intelligence into first year computing education is distinguishing between supportive and substitutive use. Supportive use enhances understanding, scaffolds reasoning, and promotes independent learning over time. Substitutive use, by contrast, replaces the thinking process entirely by generating solutions that students accept without comprehension. This distinction is particularly critical in first year contexts because foundational reasoning skills, problem solving strategies, and disciplinary thinking habits are still forming.

When AI crosses the boundary into substitution, learners may appear to complete tasks successfully while failing to develop the conceptual structures required for future learning. This creates a fragile form of competence that collapses when AI support is removed or when learners encounter novel problems that require transfer of understanding. In computing education, where later topics build cumulatively on early concepts, such gaps can have long lasting consequences.

To address this risk, ethical AI scaffolding must be grounded in clear principles that define acceptable and effective use. This paper proposes three criteria for distinguishing support from substitution in first year computing education. First, AI use should prioritise explanation over generation. Systems should be designed to clarify concepts, explore reasoning pathways, and illuminate relationships rather than produce final artefacts or complete solutions. Second, AI interactions should encourage reflection and justification rather than passive acceptance of output. Learners should be prompted to explain why an approach is valid, how a solution works, and what assumptions underpin their reasoning. Third, AI scaffolding must align explicitly with assessed learning outcomes. If assessments require conceptual explanation, reasoning, or personalised application, AI support should be structured to reinforce those demands rather than circumvent them.

Figure 1 presents a conceptual model of artificial intelligence functioning as a cognitive scaffold within first year computing education. The model integrates cognitive load theory with scaffolding principles to illustrate how AI can reduce extraneous cognitive load while preserving germane cognitive effort required for schema construction and transfer. Supportive uses of AI are positioned above the scaffold boundary and include explanation, clarification, planning prompts, reasoning checks, and verification guidance. These forms of assistance are designed to support learner cognition without displacing the core reasoning processes required for conceptual development. Below the scaffold boundary, the figure identifies substitutive uses of AI that risk undermining learning, including complete solution generation and unverified artefact production. These practices shift cognitive responsibility away from the learner and are associated with risks such as overreliance, illusions of competence, and loss of agency. The

Journal of Informatics Education and Research
ISSN: 1526-4726
Vol 6 Issue 1 (2026)

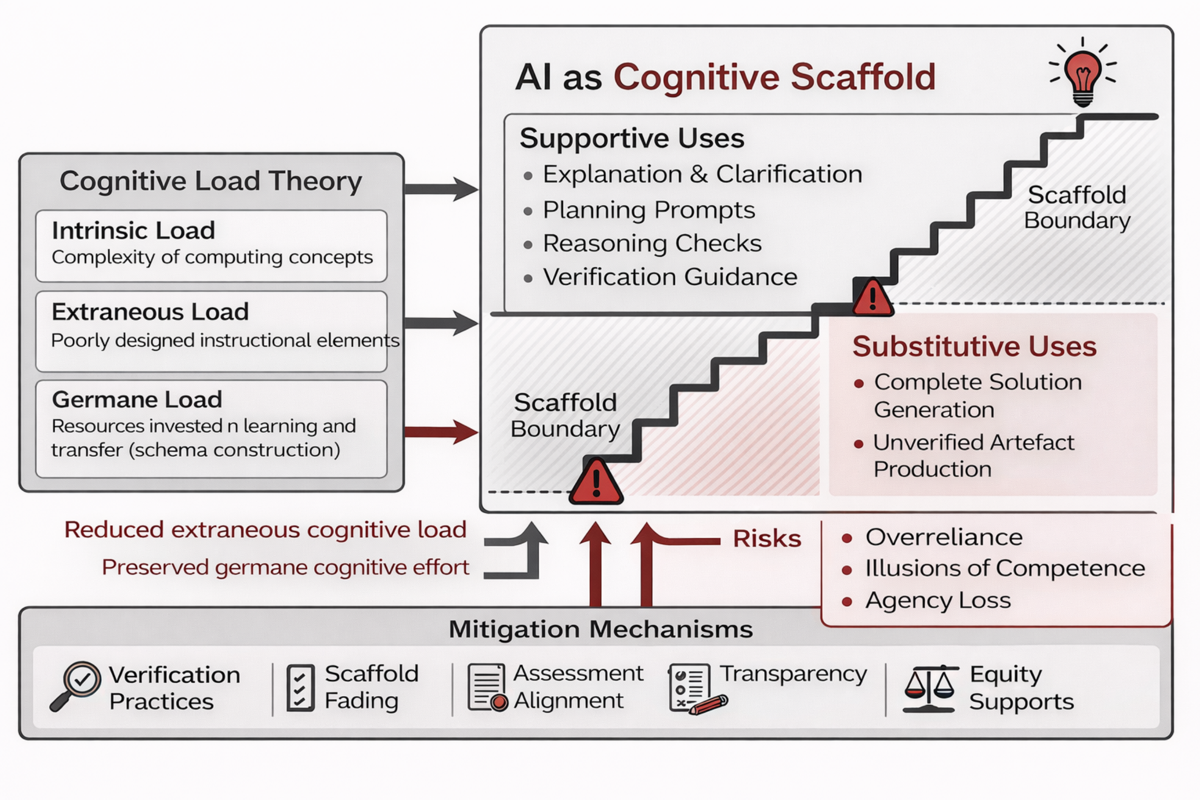



Figure 1: Artificial intelligence as a cognitive scaffold in first year computing education. The figure illustrates how AI can reduce extraneous cognitive load while preserving germane cognitive effort through supportive uses such as explanation, planning prompts, reasoning checks, and verification guidance. The scaffold boundary separates pedagogically aligned support from substitutive uses, including complete solution generation and unverified artefact production. Risk factors and mitigation mechanisms are shown to emphasise the conditions under which AI supports learning without undermining conceptual development or learner agency.

model therefore emphasises the importance of mitigation mechanisms, including verification practices, scaffold fading, assessment alignment, transparency, and equity supports. Together, these mechanisms stabilise the role of AI as a temporary and contingent support structure rather than a replacement for learner cognition, reinforcing the central argument that ethical AI integration depends on preserving student reasoning as the primary locus of problem solving.

Establishing and maintaining these boundaries requires deliberate instructional design rather than ad hoc adoption of AI tools. Educators must consider not only what AI can do, but what it should do at different stages of learning. In early stages, higher levels of explanatory support may be appropriate, while later stages should involve reduced guidance to promote independence. By clearly defining the role of AI as a temporary cognitive scaffold rather than a permanent problem solver, first year computing education can harness the benefits of artificial intelligence while safeguarding the development of durable conceptual understanding.

While the conceptual framing of artificial intelligence as a cognitive scaffold is theoretically robust, its educational value ultimately depends on how it is operationalised within real first year computing units. Without explicit instructional structures, even well intentioned AI use can drift toward substitution rather than support. This section outlines practical implementation patterns that translate the scaffolding framework into teachable, assessable, and auditable learning designs suitable for introductory computing contexts.

One effective operational model is the guided inquiry interaction pattern, where AI use is restricted to predefined categories of assistance. Students are encouraged to use AI for conceptual clarification, vocabulary explanation, reasoning checks, and planning support, while explicitly discouraged from requesting full solutions or executable artefacts. This can be reinforced through prompt exemplars provided in learning management systems, demonstrating acceptable question forms such as asking why a particular control structure is appropriate, how a concept relates to earlier material, or what assumptions underlie a given approach. By shaping how students interact with AI, educators can preserve the learners role as the primary problem solver while still reducing unnecessary cognitive friction.

A second operational pattern involves staged task design that integrates AI use differently across phases of a learning activity. Early phases may allow richer AI supported exploration, such as conceptual explanation or decomposition guidance, while later phases require students to complete synthesis, justification, or extension tasks independently. For example, students might use AI to help interpret a problem statement or identify relevant concepts, but must then produce their own algorithm design, reasoning trace, or explanation of behaviour without AI assistance. This mirrors traditional scaffold fading and helps students internalise the reasoning processes initially supported by AI.

A third pattern focuses on embedding verification as a required learning outcome. Rather than treating AI output as authoritative, students are asked to test, trace, or critique AI generated explanations. Activities can require students to identify potential edge cases, predict system behaviour under modified conditions, or explain why an AI suggested approach may fail in a particular scenario. This not only mitigates blind trust in AI, but also strengthens debugging discipline and epistemic judgement, both of which are central to computing expertise.

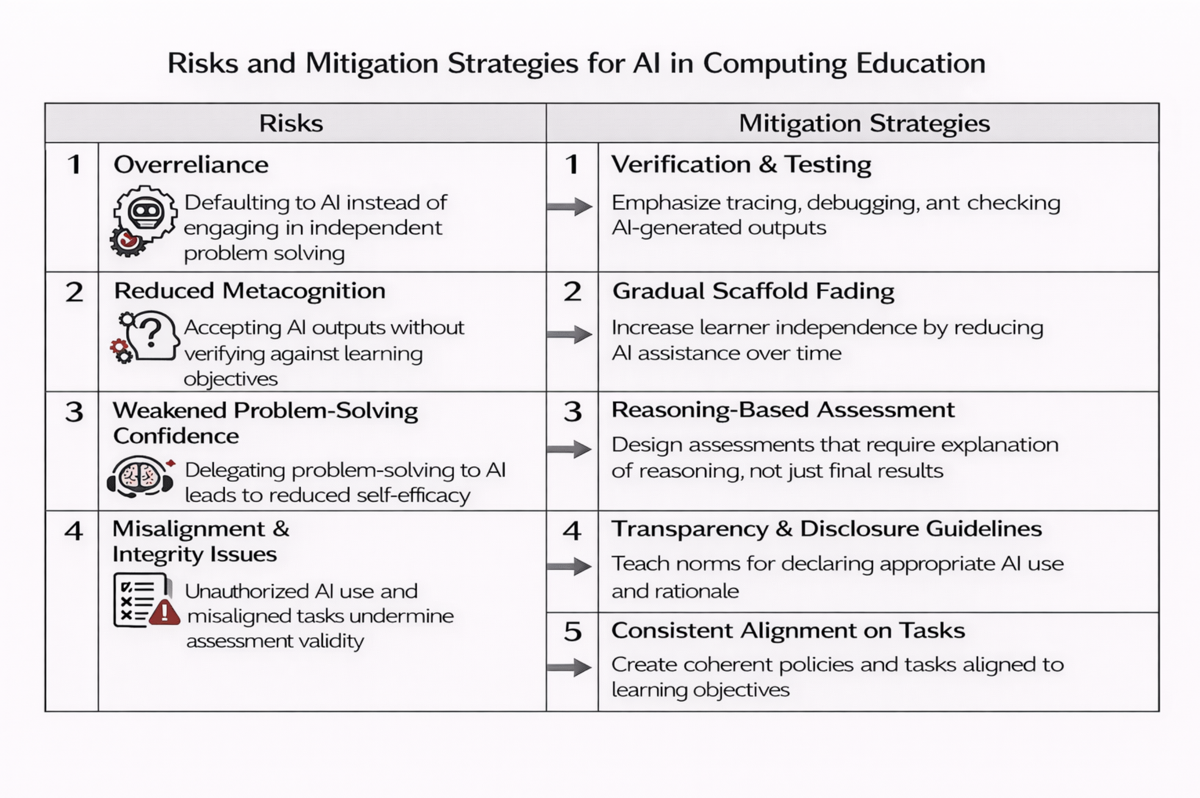
Instructor mediated reflection is also critical. Short reflective prompts can be embedded in laboratories or assessments asking students to describe how AI was used, what insights it provided, and what reasoning they ultimately adopted or rejected. These reflections reinforce metacognitive awareness and provide educators with visibility into student thinking processes. Importantly, reflection tasks should focus on reasoning development rather than compliance, framing AI disclosure as a learning activity rather than a policing mechanism.

To move beyond anecdotal claims, it is important to identify indicators that suggest AI is functioning as a scaffold rather than a substitute. One indicator is improved conceptual articulation without corresponding increases in surface level performance inflation. For example, students may demonstrate stronger explanations of control flow, data abstraction, or algorithm behaviour even when assessments are designed to minimise the value of direct solution generation. Another indicator is increased persistence on complex tasks, where students continue engaging with problems after initial difficulty rather than abandoning effort or defaulting to solution copying.

A further indicator is the quality of student questions posed to AI. When scaffolding is effective, student prompts tend to shift from outcome oriented requests toward reasoning oriented inquiries. This includes asking about why an approach works, how concepts relate, or what trade offs exist. Such prompt evolution reflects developing disciplinary thinking and increased metacognitive control.

The integration of artificial intelligence tools into computing education introduces several pedagogical risks that may influence how students develop computational thinking, reasoning ability, and independent problem-solving skills, as illustrated in Figure **??**. The framework presented in the figure identifies several major risks associated with the increasing use of AI systems in educational environments and links them with instructional strategies designed to mitigate these issues while preserving meaningful learning outcomes.

One of the most significant concerns is student overreliance on AI systems. When learners can instantly generate code, explanations, or solutions through AI tools, there is a tendency for them to default to automated outputs instead of engaging in the analytical processes required to understand the problem. In computing education, the development of algorithmic thinking, debugging ability, and systematic reasoning is essential. If these processes are bypassed, students



Risks and Mitigation Strategies for AI in Computing Education

| | Risks | | Mitigation Strategies |
|---|---|---|---|
| 1 | **Overreliance** Defaulting to AI instead of engaging in independent problem solving | 1 | **Verification & Testing** Emphasize tracing, debugging, ant checking AI-generated outputs |
| 2 | **Reduced Metacognition** Accepting AI outputs without verifying against learning objectives | 2 | **Gradual Scaffold Fading** Increase learner independence by reducing AI assistance over time |
| 3 | **Weakened Problem-Solving Confidence** Delegating problem-solving to AI leads to reduced self-efficacy | 3 | **Reasoning-Based Assessment** Design assessments that require explanation of reasoning, not just final results |
| 4 | **Misalignment & Integrity Issues** Unauthorized AI use and misaligned tasks undermine assessment validity | 4 | **Transparency & Disclosure Guidelines** Teach norms for declaring appropriate AI use and rationale |
| | | 5 | **Consistent Alignment on Tasks** Create coherent policies and tasks aligned to learning objectives |

Figure 2: Risks and mitigation strategies for the use of artificial intelligence in computing education. The figure highlights four key risks: (1) Overreliance, where students default to AI instead of engaging in independent problem solving; (2) Reduced metacognition, where learners accept AI outputs without verifying them against learning objectives; (3) Weakened problem-solving confidence resulting from delegating reasoning tasks to AI systems; and (4) Misalignment and integrity issues caused by unauthorized AI use or poorly aligned assessment tasks. Corresponding mitigation strategies include verification and testing of AI-generated outputs through tracing and debugging, gradual scaffold fading to increase learner independence by reducing AI assistance over time, reasoning-based assessments that prioritize explanation rather than only final results, transparency and disclosure guidelines that establish norms for declaring AI use, and consistent alignment of tasks with learning objectives.

may complete tasks without actually developing the conceptual understanding that the task is intended to promote. For this reason, verification and testing of AI-generated outputs becomes an important mitigation strategy. Encouraging students to trace program logic, test outputs, and debug generated solutions helps reintroduce critical evaluation into the learning process. Another important risk is reduced metacognitive engagement. Metacognition refers to the learner's capacity to reflect on their own thinking, assess whether a solution is correct, and determine whether their reasoning aligns with the objectives of the task. When AI-generated responses are accepted without scrutiny, this reflective process may be weakened. Students may rely on answers without fully understanding why those answers are correct. To address this issue, gradual scaffold fading is recommended. In this instructional approach, AI tools may initially support learning by providing guidance or examples, but the level of assistance is progressively reduced as students develop competence. This process encourages learners to become more independent and to rely increasingly on their own reasoning abilities.

A further concern involves weakened problem-solving confidence. When students repeatedly depend on AI systems to solve complex tasks, they may begin to perceive difficult problems as tasks that should be delegated to machines rather than challenges that they themselves can

Journal of Informatics Education and Research
ISSN: 1526-4726
Vol 6 Issue 1 (2026)

solve. This can reduce their confidence in their own analytical abilities and discourage deeper engagement with problem solving. To mitigate this effect, reasoning-based assessment is proposed. Instead of evaluating only the final outcome of a task, assessments can require students to explain the reasoning, logic, and decision-making steps behind their solutions. By emphasizing explanation and justification, educators reinforce the importance of human reasoning in the learning process. Issues related to misalignment and academic integrity may arise when AI tools are used without clear guidelines or without alignment to learning objectives. Unauthorized use of AI-generated solutions may undermine the validity of assessment tasks and make it difficult to determine whether a student's work reflects genuine understanding. Establishing transparency and disclosure guidelines helps address this challenge by encouraging students to clearly indicate when AI tools have been used and how they contributed to the final work. In addition, ensuring consistent alignment between tasks and learning objectives helps educators design activities where AI tools support learning rather than replace the intellectual effort required from students.

Longitudinal indicators are particularly important in computing education. If AI scaffolding is effective, students should demonstrate transfer of foundational concepts to later units where AI support may be reduced or constrained. This can be observed through improved performance on novel problem types, stronger debugging strategies, and greater confidence in independent planning and reasoning. In contrast, substitution oriented AI use often results in brittle knowledge that fails to transfer beyond familiar templates.

While this paper presents a theoretically grounded framework, empirical validation remains essential. Future studies should examine comparative learning outcomes between scaffolded AI use, unrestricted AI use, and non AI supported instruction in first year computing. Key dependent variables include conceptual mastery, cognitive load distribution, metacognitive accuracy, and retention into subsequent units. Mixed methods approaches combining performance data, prompt analysis, and student reflection can provide richer insight into how AI scaffolding operates in practice.

Experimental designs can also explore scaffold withdrawal strategies, investigating how and when reducing AI support affects learner independence. This includes identifying thresholds where support transitions from beneficial to inhibitive, as well as discipline specific differences across programming paradigms or conceptual domains. Equity focused research is also required to understand how AI scaffolding interacts with language background, prior educational experience, and confidence, ensuring that AI integration does not unintentionally amplify existing disparities.

By grounding implementation in theory, aligning design with assessment, and evaluating outcomes empirically, artificial intelligence can be positioned not as a disruptive force, but as a structured cognitive scaffold that strengthens first year computing education in a sustainable and pedagogically defensible manner.

## 3 Design Principles for Ethical AI Scaffolding

Integrating artificial intelligence into first year computing units requires design principles that treat AI as a deliberate cognitive scaffold rather than a general purpose productivity accelerator. The central objective of ethical AI scaffolding is to strengthen conceptual understanding, develop independent reasoning capacity, and support the gradual transition from guided learning to autonomous competence. This requires alignment between learning outcomes, instructional activities, assessment mechanisms, and the constraints placed on AI use. Ethical scaffolding is therefore not achieved through policy statements alone, but through coherent learning design that makes the intended role of AI explicit, visible, and pedagogically defensible.

A foundational principle is alignment with conceptual learning objectives rather than artefact production. In first year computing, the enduring outcomes are computational thinking, algorithmic reasoning, abstraction, and the ability to explain and justify behaviour. When AI is integrated without regard to these outcomes, it can shift student effort toward producing plausible outputs while bypassing the reasoning processes that units are intended to cultivate. Ethical scaffolding requires that AI interactions reinforce conceptual goals by supporting explanation, clarification, and planning rather than delivering complete solutions. Where AI generated examples are used, they should be framed as illustrative material subject to critique rather than as authoritative answers. This ensures that students engage with underlying principles rather than merely replicating patterns.

Interaction design is a second critical principle. Many of the risks associated with AI emerge from interaction patterns that reward short prompts and deliver final answers with minimal cognitive engagement. Ethical AI scaffolding requires deliberate shaping of how students interact with AI systems. Prompt exemplars should encourage reasoning oriented dialogue, such as asking why an approach works, how a concept relates to prerequisite material, or what assumptions underlie a particular method. Students should be guided to request explanations at an appropriate level of abstraction and to ask follow up questions that deepen understanding rather than terminate inquiry. By structuring interactions in this way, AI becomes a conversational scaffold that supports thinking rather than a mechanism for answer retrieval.

A third principle concerns the calibration of support. Scaffolding is effective only when it is contingent on learner needs and gradually withdrawn as competence increases. In first year computing, learners often require substantial guidance early in a unit to overcome initial barriers related to unfamiliar syntax, abstract representations, and new problem solving paradigms. AI can provide this early support by clarifying concepts and guiding decomposition. However, ethical scaffolding requires that this level of support is not maintained indefinitely. As learners gain confidence, AI interactions should shift toward higher level prompts that require students to propose solutions, defend reasoning, and identify and correct misconceptions independently. This progression mirrors traditional scaffold fading and helps prevent long term dependence on external assistance.

Assessment alignment is a fourth principle that is essential for ethical integration. If assessments reward only correct outputs, students are incentivised to use AI as a substitute for learning regardless of stated policies. Ethical scaffolding therefore requires assessment designs that elicit evidence of understanding rather than mere completion. This can include requiring students to explain program behaviour, justify design decisions, analyse trade offs, or demonstrate verification practices such as tracing and testing. When assessments value reasoning quality and conceptual articulation, AI becomes less effective as a shortcut and more effective as a preparatory learning aid. Assessment alignment also strengthens validity by ensuring that grades reflect the constructs the unit intends to measure.

Figure 3 illustrates the ethical boundaries that distinguish supportive uses of artificial intelligence from substitutive interactions that risk undermining learning in first year computing education. The model conceptualises AI as a cognitive scaffold that should assist students in understanding concepts, planning approaches, and verifying reasoning without replacing the cognitive processes required for independent problem solving. Above the scaffold boundary, AI interactions function as pedagogically aligned supports. These include explaining computing concepts in alternative forms, providing reasoning prompts that guide students through problem decomposition, suggesting next steps when learners encounter difficulty, and assisting with debugging by prompting verification and reflection. Such interactions are designed to strengthen conceptual understanding while maintaining the learner's responsibility for synthesis and decision making. Below the scaffold boundary, the figure identifies forms of AI use that shift cognitive responsibility away from the learner and therefore risk substituting for genuine learning processes. These substitutive interactions include requesting complete solutions, generating ready-made code artefacts, or automatically producing answers that students submit without verification. When such practices occur, students may appear to complete tasks successfully

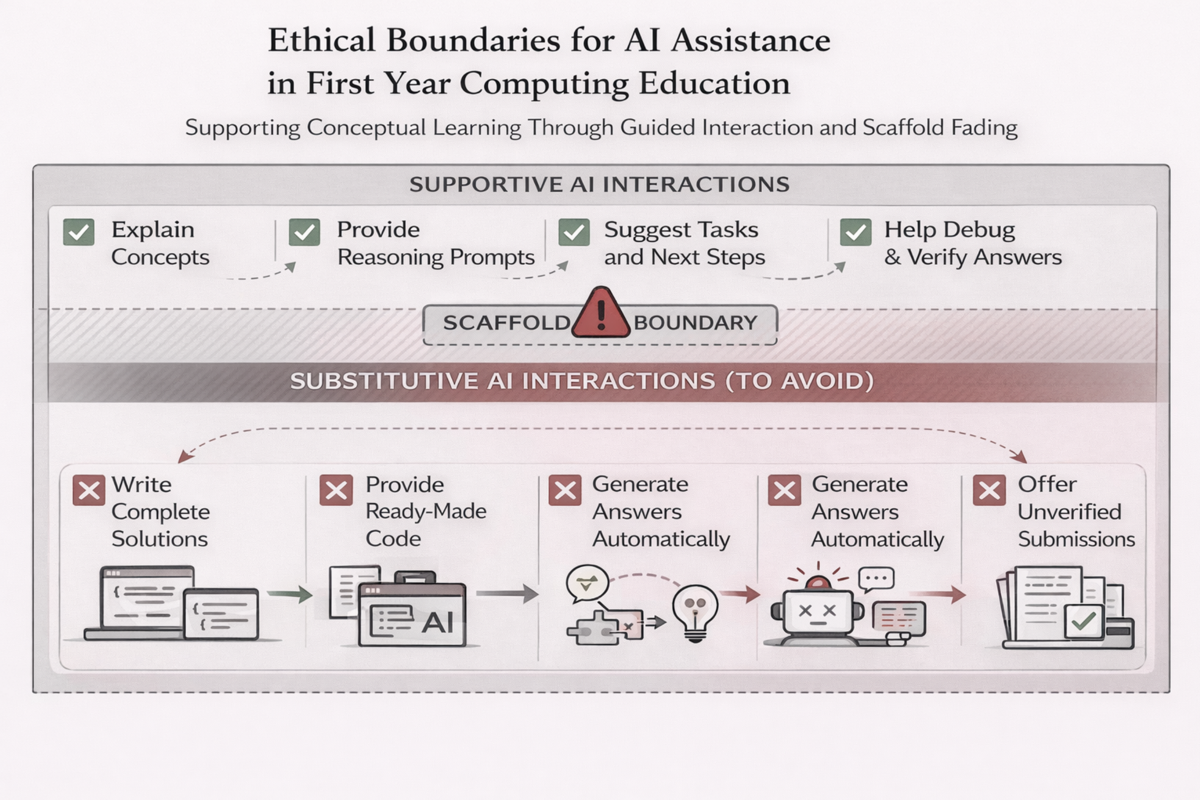



Figure 3: Ethical boundaries for artificial intelligence assistance in first year computing education. The diagram distinguishes between supportive AI interactions that enhance conceptual learning and substitutive interactions that risk replacing student reasoning. Supportive uses include concept explanation, reasoning prompts, task planning, and debugging assistance that guide learners without removing the need for independent thinking. The scaffold boundary represents the pedagogical threshold where AI assistance shifts from cognitive support to cognitive substitution. Activities below this boundary, such as generating complete solutions or producing unverified artefacts, may undermine conceptual development and should be avoided. The model reinforces the principle that AI should function as a temporary scaffold that supports learning processes while preserving student agency and reasoning responsibility.

while failing to develop the underlying reasoning skills required for algorithmic thinking, debugging, and abstraction. This creates a fragile form of competence that does not transfer effectively to novel problems or later computing courses.

The boundary depicted in Figure 3 therefore represents a pedagogical threshold rather than a technological limitation. Artificial intelligence tools can support or undermine learning depending on how they are used within instructional contexts. When interactions emphasise explanation, reflection, and verification, AI operates as a scaffold that reduces unnecessary cognitive friction while preserving germane learning effort. When interactions prioritise automated answer generation, however, the scaffold becomes substitutive and the learner's cognitive engagement is diminished. The framework highlights the importance of guiding students toward reasoning-oriented AI use so that artificial intelligence supports conceptual development rather than replacing it.

Transparency and auditability form a fifth principle. Students should be encouraged to disclose how AI was used, for what purposes, and at what stages of their problem solving process. Normalising disclosure reduces concealment behaviours and supports ethical engagement by framing AI use as a learning choice rather than a transgression. Disclosure practices can also promote metacognitive reflection, as students must articulate what assistance was helpful, what was rejected, and how final decisions were made. For educators, transparency provides insight into student learning processes and can inform instructional refinement. Importantly, transparency mechanisms should be lightweight and focused on reasoning development rather than surveillance.

A sixth principle is explicit boundary setting between support and substitution. Ethical AI scaffolding requires clear guidance on what constitutes acceptable use within a unit. This guidance should move beyond generic statements and provide concrete examples of encouraged and discouraged behaviours. Encouraged uses may include requesting conceptual explanations, seeking clarification of terminology, checking reasoning steps, or asking for hints about decomposition. Discouraged uses may include requesting complete solutions, directly reusable code, or answers that bypass required reasoning. Clear boundaries reduce uncertainty for students, support consistent marking, and align AI use with learning objectives.

Equity considerations introduce a seventh principle. AI scaffolding can exacerbate disparities if access to tools, devices, connectivity, or prompt literacy varies across students. Ethical design requires that AI is not the sole or primary scaffold on which learners depend. Units should maintain strong non AI supports, including tutor guidance, structured worksheets, worked examples, and peer learning opportunities. Prompt literacy should be taught as an academic skill aligned to learning goals, ensuring that all students understand how to request explanation and verification rather than answers. By embedding equity considerations into design, AI scaffolding can support inclusion rather than amplifying existing gaps.

An eighth principle concerns epistemic responsibility. AI systems can generate fluent explanations that appear authoritative even when incomplete or incorrect. Ethical scaffolding requires that students are taught to treat AI output as provisional and subject to evaluation. Learning activities should embed verification practices such as testing predictions, tracing execution, comparing explanations against formal definitions, and identifying failure cases. By positioning verification as a core disciplinary practice, units reinforce the idea that understanding arises from evaluation and reasoning rather than acceptance. This principle aligns AI use with professional computing practice, where outputs are routinely scrutinised through testing and review.

Consistency across units and teaching teams forms a ninth principle. Students often take multiple first year units simultaneously, and inconsistent expectations regarding AI use can create confusion and encourage strategic behaviour that undermines learning. Ethical scaffolding benefits from coherent messaging across programs, even when specific implementations vary by unit. Shared language around support versus substitution, consistent disclosure expectations, and aligned assessment philosophies help students internalise ethical AI use as a normal part of academic practice rather than as a series of ad hoc rules.

Finally, ethical AI scaffolding requires ongoing evaluation and refinement. Learning design should be informed by evidence about how students actually interact with AI and how those interactions affect learning outcomes. Educators should monitor indicators such as changes in student questioning patterns, quality of explanations, persistence on complex tasks, and transfer of concepts to later units. Student reflections and prompt analyses can provide valuable insight into whether AI is functioning as a scaffold or a substitute. This evidence based approach ensures that design principles remain responsive to learner needs and evolving technologies.

Taken together, these design principles position artificial intelligence as a pedagogical partner that amplifies learning rather than an intellectual shortcut that replaces cognitive effort. When AI support is aligned with conceptual goals, structured toward reasoning, calibrated over time, reinforced through assessment, bounded by clear expectations, and embedded within an equitable and transparent learning environment, it can function as an ethical cognitive scaffold. Under these conditions, AI integration strengthens novice learning in first year computing education while preserving the development of independent competence that underpins long term success in the discipline.

of competence, where confidence increases while diagnostic accuracy remains low.

In first year computing, this risk is consequential because learners often have limited ability to detect conceptual errors. For example, a novice may accept a plausible explanation of loop behaviour without recognising that the explanation mischaracterises variable scope, termination conditions, or state updates. Similarly, a student may accept an AI generated debugging narrative without verifying it against the actual execution trace. When students rely on AI explanations as epistemic authority rather than as provisional guidance, they may accumulate misconceptions that remain latent until they fail in later tasks. The problem is not only that AI can be wrong. The deeper problem is that AI can be convincingly wrong in ways that novices cannot easily detect.

The primary mitigation is to make verification an explicit instructional objective. Students should be taught that AI output is always provisional and must be evaluated through testing, tracing, and comparison with authoritative sources such as unit materials. Verification can be embedded as a required step in activities. For example, students can be required to generate a set of test cases, predict results, run the program, and reconcile any differences. They can be required to trace the state of key variables across iterations and compare the trace to AI explanations. By operationalising verification, the unit trains epistemic vigilance and reduces passive acceptance.

Another mitigation is to require self explanation that cannot be satisfied by copying. Students can be asked to explain why a specific approach works under specific conditions, to describe how behaviour changes when an input changes, or to justify design decisions in terms of control flow and state. Short reflective prompts can ask what the student initially believed, what evidence changed that belief, and what the final understanding is. These prompts support calibration of confidence and encourage learners to monitor their own understanding more accurately.

A third risk is diminished confidence in independent problem solving, paradoxically caused by persistent access to strong assistance. While AI can initially reduce anxiety and frustration, continuous reliance can lead students to internalise the belief that meaningful progress requires external support. When a learner repeatedly experiences success only in the presence of AI, the locus of control can shift away from the learner, weakening agency and resilience. This risk is particularly relevant in introductory computing, where early setbacks are common and where students often attribute difficulty to personal inability rather than to normal learning processes.

Loss of agency has downstream consequences. Students may avoid tasks that require independent reasoning, delay engagement until they can consult AI, or disengage when AI output is restricted. In assessments or professional contexts where AI access is limited, learners may experience disproportionate stress and performance degradation. In addition, loss of agency can reduce motivation to build internal debugging discipline, since students may treat debugging as a service provided by tools rather than as a core thinking practice.

A mitigation strategy is to design activities that preserve productive struggle while preventing unproductive failure. This can be achieved through bounded hint systems, where students first attempt a problem, then access limited hints, then access AI prompts that guide reasoning rather than providing answers. The goal is to preserve the psychological experience of solving while reducing the likelihood of prolonged stuck states that lead to disengagement. Educators can explicitly frame struggle as normal, then provide scaffolds that enable progress without removing ownership.

Another mitigation is to include regular low stakes independence checks, such as short in class reasoning tasks, oral micro explanations, or trace based questions where students must explain behaviour without AI. These tasks provide feedback to both student and instructor about developing competence and reduce the risk that learners build an identity as dependent on external tools.

Generative AI can introduce misconceptions through incorrect, oversimplified, or context misaligned explanations. Even when outputs are technically correct, they may conflict with unit conventions, teaching sequences, or the intended conceptual framing. For example, an AI system may introduce advanced constructs that are not yet taught, propose patterns that bypass intended learning objectives, or present multiple solutions without signalling which aligns with the pedagogical focus of the unit. This can create conceptual noise that increases extraneous load rather than reducing it.

A related risk is inconsistency. Students may receive different explanations, different solution strategies, or different conceptual framings depending on how they prompt, what tool they use, and what language they use. This can fragment the shared learning experience in a cohort and complicate teaching, since student misconceptions may no longer arise from common sources. In large classes, this variability can overwhelm instructional support because learners arrive with heterogeneous, AI shaped mental models.

One mitigation is to provide official prompt templates aligned to unit learning outcomes. These templates can steer students toward asking for conceptual explanations at the level of the unit, requesting hints in the units style, and avoiding requests for full solutions. A second mitigation is to provide authoritative reference anchors, such as a unit glossary, standard patterns for tracing and debugging, and canonical examples. Students can be instructed to reconcile AI explanations with these anchors, reducing drift.

In some contexts, educators can embed AI prompts directly into activities, specifying what the AI should be asked and what the student must produce afterwards. This turns AI into a controlled scaffold rather than an unbounded generator. For example, a lab can instruct students to ask AI for a conceptual explanation of a loop invariant, then require the student to write a trace for a given input and identify where the invariant holds. This approach reduces uncontrolled variation while preserving adaptive explanation.

Integrity related risks include unacknowledged substitution of work, direct generation of assessed artefacts, and increased difficulty in validating authorship. However, beyond policy compliance, a deeper educational risk is construct irrelevance, where assessments inadvertently measure access to tools or prompt skill rather than the intended learning outcomes. If an assessment rewards only the production of correct code outside controlled conditions, then AI can dominate performance, reducing the validity of grades as indicators of competence. This can also reduce perceived fairness among students and weaken trust in assessment systems.

Integrity risks are amplified when expectations are unclear, inconsistent across units, or framed only as prohibitions. Students may not understand what is acceptable, may interpret rules differently, or may rationalise substitution as normal practice if they perceive that outcomes matter more than learning. In first year cohorts, ambiguity is particularly harmful because novice students are still learning academic norms and may be more influenced by peer narratives than by policy documents.

The strongest mitigation is to redesign assessments so that they require evidence of reasoning that is difficult to outsource. This includes requiring explanation of behaviour, justification of decisions, analysis of trade offs, and demonstration of verification practice. Process artefacts can be included, such as annotated traces, test design rationales, and debugging narratives grounded in observed outputs. Where code is assessed, tasks can include personalised parameters, controlled variations, or constraints that require adaptive reasoning rather than direct reuse.

Another mitigation is to normalise disclosure of AI use. Students can be required to include a brief AI use statement describing what AI assisted with, what was accepted, what was rejected, and what verification was performed. In selected tasks, students can submit prompt logs or summaries of interactions. The goal is not surveillance, but the creation of a reflective habit that frames AI use as part of learning practice. Disclosure also reduces concealment incentives and supports consistent marking expectations.

Equity risks arise when students have differential access to AI tools, different levels of digital

Journal of Informatics Education and Research
ISSN: 1526-4726
Vol 6 Issue 1 (2026)

literacy, or different capacity to craft effective prompts. Even when tools are broadly available, differences in device quality, connectivity, language proficiency, and prior educational experience can shape how effectively learners use AI support. Students who can prompt well may receive better guidance, while students who struggle to articulate questions may receive low quality assistance. This can widen gaps that already exist in first year cohorts, particularly in diverse intakes where students vary widely in prior exposure to computing and academic communication.

Equity issues also include cultural and linguistic factors. Students working in a second language may have difficulty expressing misconceptions precisely, leading to less effective AI responses. Students may also be more likely to accept AI output without critique if they lack confidence in their own interpretation of technical language. In addition, some students may be ethically cautious and avoid AI, while others exploit it aggressively, creating uneven learning conditions.

A mitigation strategy is to provide consistent access pathways and to teach prompt literacy as an academic skill. Prompt literacy in this context should not be framed as a productivity trick. It should be framed as a method for requesting conceptual explanation, eliciting reasoning rather than answers, and prompting for verification steps. Teaching prompt literacy reduces differential advantage and aligns tool use with learning goals. Provisioning can include guidance on accessible tools, availability of institutional resources, and clear alternatives for students who cannot or do not wish to use AI.

Equity is also supported by maintaining strong non AI scaffolds, including tutor consultations, peer learning structures, guided worksheets, and high quality exemplars. AI should be one scaffold among many, not the only scaffold. Inclusive design also suggests providing question stems, glossaries, and structured debugging checklists that reduce reliance on sophisticated prompting.

Another risk involves privacy and data handling. Students may paste assessment prompts, code, or personal information into third party systems without understanding data retention, model training, or sharing risks. They may also share proprietary materials unintentionally. Instructors may also face boundary confusion regarding what kinds of data can be processed through AI tools, particularly when student work includes identifying information or when assessment tasks are confidential.

Mitigation requires explicit guidance on what may be shared with AI tools and what must not be shared. Students should receive simple rules that prohibit uploading personal data, confidential assessment content, or institutional documents to unapproved systems. Safe prompting practices can be taught, including how to abstract a problem and describe it without sharing sensitive content. Institutions can support this by providing approved tools or by publishing clear guidance consistent with privacy obligations.

A final risk is pedagogical drift, where AI becomes integrated as a convenience tool rather than as a scaffold aligned to learning objectives. Drift can occur when staff adopt AI informally without redesigning activities and assessments, or when students use AI in ways that bypass intended conceptual trajectories. Over time, the unit may shift from developing reasoning to producing artefacts, not by design but by inertia. This drift can be difficult to detect if superficial performance appears strong while conceptual depth weakens.

Mitigation requires systematic alignment between learning outcomes, activities, assessment criteria, and AI use expectations. Educators can implement periodic checks such as reviewing common AI prompts students report using, analysing where misconceptions cluster, and adjusting scaffolds accordingly. Unit policies can specify the intended role of AI at each stage of the unit and can be reinforced through learning activities that explicitly model ethical use. Professional development for teaching staff can also support consistent implementation across teaching teams, reducing conflicting messages.

Effective mitigation is multi level. At the student level, mitigation emphasises verification habits, self explanation, and reflective disclosure. At the activity level, mitigation emphasises calibrated scaffolds, bounded hints, and staged task design with scaffold fading. At the assessment level, mitigation emphasises construct validity by prioritising reasoning evidence and adaptive problem solving. At the institutional level, mitigation emphasises consistent policy, equitable access, privacy guidance, and staff development. The objective is not to eliminate AI, but to stabilise its role as a cognitive scaffold that reduces extraneous burden while preserving the cognitive work required for durable competence.

When these mitigation strategies are implemented coherently, the risks associated with AI become manageable within a principled learning design. Overreliance is countered by forced cognitive contribution and scaffold fading. Illusions of competence are countered by verification and metacognitive prompting. Agency loss is countered by structured productive struggle and independence demonstrations. Misconceptions are countered by alignment anchors and curated prompts. Integrity and construct risks are countered by assessment redesign and transparency norms. Equity risks are countered by access provisioning and prompt literacy teaching. Privacy risks are countered by explicit data rules. Pedagogical drift is countered by alignment checks and quality assurance. Under these conditions, artificial intelligence can be integrated into first year computing in a way that strengthens learning while preserving ethical standards and educational validity.

## 5 Implications for Computing Education

Integrating artificial intelligence as a cognitive scaffold has significant implications for curriculum design, assessment, teaching practice, and institutional governance within first year computing education. At the curriculum level, units can no longer assume that learners work primarily from static resources and delayed human feedback. AI assistance is now part of the learning environment, whether formally endorsed or informally used, and this shifts the baseline conditions under which novice learners engage with computing concepts. The implication is that first year curricula should be designed with explicit intent regarding where AI can reduce extraneous burden and where students must still engage in core cognitive work, such as decomposition, tracing, verification, and synthesis. A curriculum that ignores AI availability risks misalignment between learning outcomes and student practices, while a curriculum that embraces AI without boundaries risks producing fragile competence that does not transfer to later units.

At the level of learning outcomes, the presence of AI increases the importance of outcomes that capture conceptual mastery and reasoning quality rather than mere artefact production. In introductory programming, the enduring competencies are the ability to explain program behaviour, reason about state and control flow, debug systematically, and apply abstractions across contexts. If AI can readily produce plausible code, then the educational value shifts toward what learners can justify, predict, and verify. This implies that outcomes should explicitly include epistemic judgement, such as the ability to evaluate the correctness of explanations, to detect inconsistencies between predicted and observed behaviour, and to validate solutions against constraints. When framed in this way, AI becomes a tool that can support learning without redefining competence as tool assisted output.

Curriculum sequencing also changes. In many introductory courses, conceptual progression is designed under the assumption that students build mental models gradually through repeated practice with feedback that is often limited. AI can accelerate access to feedback and alternative explanations, which may allow faster progression through some content, but it can also mask gaps that would otherwise surface through struggle. The implication is that curriculum designers should integrate deliberate checkpoints that reveal whether conceptual structures have formed. These checkpoints can include trace based reasoning tasks, prediction tasks where students anticipate outputs before running code, and explanation tasks that require learners to articulate why a method works under specific conditions. Such tasks can be low stakes and frequent, serving as diagnostic mechanisms that stabilise learning trajectories in the presence of abundant AI assistance.

Another implication concerns the design of learning activities. Laboratory tasks and tutorials can incorporate AI as one scaffold among many, but they should avoid treating AI as a universal shortcut. Activities should be designed to make the student the primary agent of reasoning, with AI supporting clarification and guided inquiry. For example, tasks can require students to produce an initial plan, consult AI for conceptual clarification, then revise the plan and justify the revision. Tasks can also require students to generate test cases and use AI output only as a hypothesis to be evaluated. In this model, AI is positioned as a conversational tutor that supports sense making while the learner remains responsible for constructing and defending a solution.

Teaching practice is also affected. Educators shift from being primarily content deliverers to being learning architects who design environments where students build durable competence while AI is available. This includes teaching verification habits explicitly, modelling how to ask reasoning oriented questions, and demonstrating how to critique and validate AI outputs. Educators also take on the role of norm setters who communicate boundaries clearly. Instead of generic statements about AI being allowed or not allowed, teaching practice benefits from concrete examples of acceptable use, such as asking for an explanation of why a loop terminates, requesting a conceptual analogy for variable scope, or seeking hints about decomposition. Equally, educators should provide examples of unacceptable use, such as requesting complete solutions or directly reusable code for assessed tasks, when those behaviours undermine learning outcomes.

Feedback systems within units must also adapt. Traditional feedback in first year computing often relies on automated testing frameworks and tutor marking, which can emphasise correctness and completion. In a scaffolded AI environment, feedback should be rebalanced toward reasoning quality. This includes providing feedback on conceptual explanations, on the quality of test case design, on the completeness of traces, and on the coherence of justification. Automated tools can support this by prompting students to provide explanations alongside code submissions, requiring students to describe how they verified correctness, and flagging patterns consistent with superficial completion. The aim is to create a feedback ecology that reinforces learning processes, not just end products.

Assessment design becomes a central lever for ensuring that AI supports rather than replaces learning. If assessments reward only correct final outputs, students will rationally use AI for substitution, because the incentive structure favours completion over understanding. The implication is that assessments should be designed so that the construct being measured is conceptual mastery and reasoning, not merely the ability to produce code. This can be achieved by requiring explanation of behaviour under specified inputs, justification of design choices, analysis of trade offs, and demonstration of verification practices. Assessments can include personalised parameters, controlled variations, and constraints that require adaptive reasoning rather than direct reuse. Where take home assessments remain appropriate, they can be paired with in class reasoning checks, oral micro explanations, or trace based tasks that provide complementary evidence of competence.

Validity and fairness considerations become more complex. When AI is available, output alone becomes weaker evidence of student competence. This challenges the validity of some traditional assessment formats and requires a shift toward multi evidence models where reasoning, explanation, and verification are foregrounded. Fairness also requires consistent expectations across teaching teams and across units. If one unit permits broad AI use and another prohibits it, students may receive conflicting messages and adopt strategies that undermine learning. Institutions should therefore aim for coherence in AI policy application, while still allowing disciplinary nuance and unit specific design decisions.

Academic integrity also changes in character. Integrity concerns remain, but they should be reframed as a design problem as much as a compliance problem. If a task is easily solvable through direct AI generation, then the task may be measuring the wrong construct in an AI rich environment. The implication is that integrity is strengthened when tasks require personalised reasoning and when disclosure is normalised. Requiring brief AI use statements can support transparency while also encouraging reflection about what was accepted, what was rejected, and how verification was performed. When disclosure is treated as a learning practice rather than a policing mechanism, students are more likely to engage ethically and instructors gain insight into where learners are struggling.

Institutional governance has further implications. Institutions need unit level AI use statements that specify boundaries in concrete terms and align those boundaries with learning outcomes. Governance also includes privacy guidance, since students may not understand what kinds of data are safe to share with external tools. Institutions should provide clear rules about avoiding personal data, confidential assessment content, and proprietary materials in unapproved systems. Where possible, institutions can support safer usage through approved platforms or through guidance that teaches students how to abstract problems without exposing sensitive content.

Equity must also be addressed systematically. AI scaffolding can widen disparities if students have differential access to tools, devices, connectivity, or prompt literacy. The implication is that first year computing units should teach prompt literacy as an academic skill aligned to learning goals, emphasising how to request conceptual explanation, how to elicit reasoning rather than answers, and how to prompt for verification steps. Units should also maintain strong non AI scaffolds, such as tutor consultations, guided worksheets, structured debugging checklists, and canonical examples. AI should be one support channel among many, not the sole support on which learners depend.

The integration of AI as a scaffold also affects how cohorts develop a shared language and shared practices. In traditional settings, students converge on common conceptual framings presented by instructors and textbooks. With AI, learners may receive heterogeneous framings that can fragment cohort understanding. The implication is that units should provide authoritative anchors, including consistent terminology, standard tracing methods, and canonical examples that students can use to reconcile AI explanations with course intent. Educators can also incorporate activities where students compare alternative explanations and evaluate which aligns with principles taught in the unit, thereby turning heterogeneity into an opportunity for critical evaluation rather than a source of confusion.

There are also implications for progression into later units. If AI scaffolding is effective, students should transfer foundational concepts and verification habits into more advanced contexts, including data structures, software engineering practices, databases, and networking. If AI scaffolding is substitutive, students may appear to succeed early but struggle later when tasks demand independent synthesis. The implication is that programs should track longitudinal indicators such as performance on novel problem types, quality of debugging strategies, ability to justify design decisions, and confidence in independent planning. These indicators can inform whether scaffold fading is appropriately timed and whether additional supports are required.

Professional development for educators becomes essential. Teaching staff need shared understanding of what constitutes support versus substitution, how to design assessments that elicit reasoning evidence, and how to teach verification practices explicitly. Professional development can also support consistent messaging to students, reducing confusion and perceived unfairness. Educators benefit from guidance on prompt patterns that promote explanation, on designing tasks resilient to direct generation, and on interpreting student work in a context where artefact quality alone is not sufficient evidence of competence.

Finally, the integration of AI scaffolding reshapes the research agenda in computing education. Units and programs need empirical evaluation of how AI scaffolds affect cognitive load, conceptual mastery, metacognitive accuracy, and retention. Mixed methods research that combines performance data, analysis of student prompts, reflective disclosures, and qualitative interviews can provide insight into how learners actually interact with AI and which designs best support learning. Research should also examine scaffold fading strategies, identifying when support transitions from beneficial to inhibitive, and how that transition differs across student backgrounds and prior experience. Equity focused research is critical to ensure that AI integration does not amplify disparities, and that alternative scaffolds remain strong for students who have limited access or who choose not to use AI.

Overall, the implications of AI as a cognitive scaffold are not confined to tool adoption. They require coordinated redesign across curriculum outcomes, activity structures, assessment validity, feedback systems, integrity norms, equity supports, privacy governance, staff capability, and program level monitoring. When these elements are aligned, AI can reduce unnecessary barriers for novice learners while strengthening the core reasoning practices that first year computing aims to develop, thereby preserving educational quality in an environment where powerful assistance is now a persistent feature of student learning.

## 6 Conclusion

Artificial intelligence has the potential to function as an effective cognitive scaffold in first year computing education when it is grounded in sound pedagogical principles and implemented through deliberate instructional and assessment design. Rather than treating AI as a binary presence that must be either embraced without constraint or prohibited to protect standards, this paper has argued for a third position in which AI is integrated as a structured support that reduces unnecessary friction while preserving the primacy of conceptual understanding. In introductory computing, the central objective is not the rapid production of artefacts but the development of durable reasoning capabilities, including the ability to plan, decompose, trace, debug, verify, and explain program behaviour across contexts. AI can support these capabilities when it is used to clarify concepts, provide alternative representations, and prompt reasoning oriented dialogue, but it can undermine them when it becomes a substitute for the cognitive work that novices must perform to form stable mental models.

A key contribution of the cognitive scaffold framing is that it treats AI assistance as temporary, contingent, and subject to withdrawal as competence increases. This framing aligns with established learning theory in which scaffolds are calibrated to learner needs and removed as independence develops. In first year computing, where intrinsic cognitive load is often high due to abstraction, symbolic reasoning, and the need to coordinate multiple interacting elements under working memory constraints, AI can be beneficial by reducing extraneous load. It can clarify task requirements, reduce confusion caused by vague instructions, and provide timely feedback that supports progress. However, reducing extraneous load is educationally valuable only when it protects cognitive capacity for germane processing, meaning effort invested in schema formation, transfer, and conceptual integration. If AI removes the need for germane effort by producing complete solutions or by enabling learners to bypass planning and verification, then it does not function as a scaffold. It functions as a replacement, producing brittle competence that fails under novel constraints and weakening the development of disciplinary thinking habits.

The distinction between support and substitution therefore becomes central. This paper has emphasised that the boundary is not defined by whether AI is used, but by what cognitive work remains with the learner. When AI interactions prioritise explanation over generation, when students are prompted to justify decisions, and when verification practices are embedded as routine, AI can strengthen learning rather than dilute it. In contrast, when learners treat AI output as authoritative, accept generated solutions without critique, or externalise core reasoning processes, the result is a degradation of metacognitive accuracy, reduced agency, and weakened transfer to later units. These outcomes are particularly damaging in computing education because early concepts are cumulative prerequisites for later content, and small gaps in mental models can propagate into major barriers as tasks increase in complexity.

The paper has also argued that effective integration depends on alignment across curriculum, learning activities, assessment design, and institutional guidance. Curriculum design must anticipate that AI is available and will be used, and therefore must foreground outcomes that remain meaningful in an AI rich environment, including explanation, reasoning quality, and epistemic judgement. Learning activities should position AI as a guided inquiry tool rather than a universal shortcut, requiring learners to contribute plans, traces, test design, and reflective explanations that demonstrate ownership of the problem solving process. Assessment design should shift away from measuring output alone and toward measuring evidence of understanding through justification, prediction, verification, and adaptive reasoning under constraints. When assessment constructs reward conceptual mastery, AI becomes less useful as a shortcut and more useful as a support for learning.

Risks remain substantial, but they are manageable when treated as predictable design challenges rather than as purely behavioural problems. Overreliance can be mitigated through scaffold fading and through activity structures that require student contribution before and after AI interaction. Illusions of competence can be mitigated by teaching verification as an explicit learning outcome and by embedding tracing, testing, and critique as required practices. Loss of agency can be mitigated through bounded support that preserves productive struggle and through frequent low stakes demonstrations of independent reasoning. Misconceptions and inconsistency can be mitigated by providing authoritative anchors, prompt exemplars aligned to unit intent, and curated patterns for explanation and debugging. Integrity and validity concerns can be mitigated by designing assessments that elicit reasoning evidence, by normalising disclosure, and by maintaining coherent policies that students can interpret consistently. Equity concerns require special attention through access provisioning, prompt literacy instruction framed as an academic skill, and the maintenance of strong non AI scaffolds that support diverse learners.

The broader implication is that the presence of AI forces a clearer articulation of what first year computing education is trying to produce. If the goal is to develop students who can reason about computation, not merely generate code, then AI integration becomes an opportunity to strengthen teaching practice by making reasoning explicit, by emphasising verification and epistemic judgement, and by designing assessments that reflect genuine competence. In this sense, AI can prompt beneficial reform, encouraging educators to move beyond artefact focused evaluation and toward richer measures of understanding that better align with professional computing practice, where correctness is established through testing, review, and reasoning rather than through superficial plausibility.

Future work should focus on empirical validation and refinement of scaffolded AI models across diverse first year cohorts. Comparative studies can examine differences between scaffolded AI use, unrestricted AI use, and non AI supported instruction, measuring outcomes such as conceptual mastery, cognitive load distribution, metacognitive calibration, persistence, and retention. Research should also examine scaffold withdrawal strategies to identify when and how reducing AI support best promotes independence without triggering disengagement. Prompt analysis and reflective disclosure can provide insight into how students actually use AI and whether interactions evolve from outcome seeking to reasoning seeking over time. Equity focused research is essential to understand how AI scaffolding interacts with language proficiency, prior educational experience, confidence, and access, ensuring that integration strategies do not unintentionally widen disparities.

In conclusion, artificial intelligence can be integrated into first year computing education in a way that strengthens learning, integrity, and educational quality, but only when it is treated as a deliberately designed cognitive scaffold rather than an unbounded productivity tool. The core requirement is not the elimination of AI, but the preservation of learner cognition as the central locus of problem solving. When AI is aligned with conceptual goals, structured toward explanation, constrained to support verification and reflection, and faded as competence develops, it can reduce unnecessary barriers for novices while reinforcing the reasoning practices that define computing expertise. Under these conditions, AI becomes not a threat to foundational learning, but a tool that can help more students build durable competence and progress with confidence into the broader computing curriculum.References

[1] Watts, K. J. (2025). Paying the cognitive debt: An experiential learning framework for integrating AI in social work education. Education Sciences, 15(10), 1304.

ards Intelligent-TPACK: An empirical study on teachers' professional
y integrate artificial intelligence-based tools into education. Computers