From Usage Intensity to Relationship Quality: A Typology of Student-AI Use and Design Propositions for Generative AI in Higher Education

Riem Yasin[1], Samir Yasin[2] and Carsten Lanquillon[3]

[1]Heilbronn University of Applied Sciences, Germany, riem.yasin@hs-heilbronn.de
[2]Heilbronn University of Applied Sciences, Germany, sayasin@stud.hs-heilbronn.de
[3]Heilbronn University of Applied Sciences, Germany, carsten.lanquillon@hs-heilbronn.de

Abstract

Competence frameworks for generative AI in higher education specify what students should be able to do, yet they leave the everyday interaction pattern that produces these abilities under-specified. This paper proposes the quality of the student-AI relationship as this missing design object. Licklider's vocabulary of symbiosis informs this typology, although it replaces his ideal of tight coupling with preserved independent capability. Two states result. Capability-returning use strengthens the student's own capability through delegation; dependent use leaves students unable to produce the delegated work alone. We hypothesise that the interface layer shapes which relationship forms. Three observable markers make the two states assessable, namely verification of AI output, task-scoped delegation, and dependence awareness. Descriptive evidence from two open survey datasets illustrates the markers. Among 218 high-frequency users in a single-institution survey, the markers separate four response combinations (29, 33, 17, and 20 %). In a cleaned global dataset, ethical concern is associated with somewhat lower intensive use, yet intensive use remains common under high concern. The contribution consolidates overreliance, assessment-design, and self-regulation findings into a relationship-quality typology and five design propositions, with functional legibility as the novel proposition.

Keywords: *AI in education; human-AI interaction; AI literacy; cognitive offloading, interaction design; student-AI relationship*

1. Introduction

Survey evidence has made ChatGPT use in higher education visible at scale. An early global study reported a sample of 23,218 students from 109 countries and territories and identified brainstorming, summarising texts, and finding research articles as common academic uses [1]. The 2024-2025 dataset used in this paper contains 22,963 responses from 120 countries and territories and records use, perceived capabilities, ethical concerns, satisfaction, study outcomes, and skills development [2]. Teachers also report that generative AI (GenAI) requires changes in teaching and assessment, including stronger attention to AI understanding, critical thinking, ethical judgement, creativity, collaboration, and learning processes [3]. These findings establish the scale of use and the pressure on higher education. They leave open which quality of student-AI interaction learning environments should cultivate.

Usage frequency provides a weak measure of student use of GenAI when it treats different learning interactions as equivalent. Two students may use the same system equally often while delegating different parts of the learning process. Some students compare an AI-generated explanation with course material, identify gaps, and revise their own draft. Others submit a model-generated answer without inspecting its reasoning or evidence. A frequency-based measure counts both cases as equal use, although they allocate cognitive work differently. The relevant object is therefore relationship quality, understood as the pattern through which students allocate, monitor, and reclaim cognitive work in interaction with AI. Interfaces, defaults, prompts, checkpoints, and course rules shape this pattern. These elements raise or lower the effort of verification, normalise particular forms of delegation, and decide when human judgement re-enters the task.

Licklider's account of man-computer symbiosis gives this shift its theoretical starting point. He framed interactive computing as a close coupling in which humans set goals, formulate hypotheses, determine criteria, and evaluate outcomes while computers perform routinisable work that prepares the way for insight and decision [4]. For GenAI in higher education, this framing turns delegation into the central design question. Risko and Gilbert define cognitive offloading as external action that changes a task's information-processing requirements and lowers cognitive demand. Their analysis also links offloading decisions to metacognitive judgements about one's own ability [5]. In AI-supported studying, delegation becomes risky when students let the system complete whole performances or accept its output without verification. This risk appears in adjacent evidence. Zhai et al. review the risks of over-reliance on AI dialogue systems for decision-making, analytical reasoning, and critical thinking [6]. Gerlich links frequent AI-tool use to lower critical-thinking scores and identifies cognitive offloading as the mediating factor [7]. The central design issue is therefore how delegation is organised, so that AI-supported work stays checked, bounded, and connected to the learner's own capability.

This paper develops a relationship-quality typology for that concern. The typology distinguishes capability-returning use, in which AI-supported delegation strengthens the student's own capability, from dependent use, in which students can no longer produce the delegated performance alone. Between these states sits the interface layer, understood as the visible and institutional arrangement of prompts, labels, defaults, verification steps, and assignment

rules. This layer links the typology to two research strands. Human-centred AI pursues high automation and high human control as a joint design goal [8], while hybrid intelligence treats human-AI work as a division of labour that draws on complementary strengths [9].

The contribution is conceptual and threefold. It consolidates prior work on overreliance, control allocation, and AI assessment into a relationship-quality typology for student-AI interaction in higher education. The typology defines capability-returning use, dependent use, and the mediating interface layer (Section 3), and makes these concepts assessable through three observable markers, namely verification of output, task-scoped delegation, and dependence awareness (Section 3.3). It then states five design propositions for AI-supported learning environments, with functional legibility marking the proposition least covered by prior assessment and overreliance work (Section 5). A descriptive illustration from two open survey datasets [2], [10] examines whether the named marker combinations appear in student populations and how ethical concern relates to frequent use (Section 4).

2. Background

Large-scale survey work defines the adoption problem. Ravšelj et al. analyse higher-education students' early perceptions of ChatGPT and report task-specific uses such as brainstorming, summarising, and academic writing [1]. The 2024-2025 follow-up dataset records usage, capabilities, regulation and ethical concerns, satisfaction, study outcomes, and skills development for 22,963 responses from 120 countries and territories [2]. Bower et al. add the teaching perspective by showing that teachers largely expected generative AI to require changes in teaching, learning outcomes, and assessment, while their proposed assessment responses varied widely [3]. Holmes et al. sharpen the design problem [11]. The authors characterise much AI in education (AIED) as reflecting or automating existing educational assumptions and practices, whereas stronger uses would open possibilities that challenge established pedagogies or help teachers work more effectively. Together, these sources establish scale, institutional pressure, and the need for design direction. They do not yet specify which qualities of student-AI interaction learning environments should cultivate.

Although competence frameworks clarify what students should be able to do with AI, they leave the everyday interaction pattern that produces these abilities under-specified. AI literacy is framed as the capacity to evaluate AI critically, communicate and collaborate with it effectively, and use it as a tool [12]. Review work organises this capacity into knowing and understanding AI, using and applying AI, evaluating and creating AI, and addressing ethical issues [13]. UNESCO turns the same competence target into twelve student competency blocks [14]. These blocks span a human-centred mindset, the ethics of AI, AI techniques and applications, and AI system design, and they advance from Understand to Apply and Create. In higher-education terms, Ehlers frames the same demand as self-organised action in open, uncertain, and digitally shaped problem situations [15]. These accounts clarify what students should become capable of. The unresolved design question is how repeated student-AI interaction must be structured so that delegation supports that capability instead of replacing the work through which it develops.

The mechanism linking delegation to learning is cognitive offloading. External action can reduce a task's information-processing demands and improve performance when internal capacity is limited, yet offloading also changes which parts of the task remain cognitively active for the learner [5]. Task demand and metacognitive self-judgements together shape offloading [5]. The same analysis names memory costs, automation complacency, automation bias, and possible skill decay as risks of technology-based offloading. In AI-supported education, this mechanism becomes critical when students accept AI-generated recommendations without questioning them. Zhai et al. define this pattern as over-reliance and review evidence linking it to risks for decision-making, critical thinking, and analytical reasoning in educational and research contexts [6]. Gerlich's mixed-method survey reports the same concern empirically, since frequent AI-tool use is negatively associated with critical-thinking scores, with cognitive offloading mediating the association [7].

Design follows from this diagnosis. Shneiderman's HCAI framework separates automation from human control and treats their combination as a design target, so that systems raise performance while preserving self-efficacy, mastery, creativity, and responsibility [8]. Dellermann et al. sharpen this into a relationship claim. Hybrid intelligence allocates work between human and artificial agents according to complementary strengths, so that the joint system achieves results neither agent could reach alone and improves through mutual learning [9]. Translated into education, this principle becomes a question of learning activity.

Student-AI collaboration is shaped by curriculum design, patterns of interaction, learning environments, and staged development [16]. In academic writing, iterative interaction with a GenAI assistant was associated with higher writing performance than linear use that treated AI primarily as a supplementary information source [17]. Together, these findings direct interfaces and courses to structure delegation through verification, explicit task boundaries, and dependence awareness.

Recent design and assessment work already addresses parts of this problem. In AI-assisted decision tasks, cognitive forcing functions reduced overreliance more effectively than simple explanations, although participants rated the stronger interventions less favourably [18]. In a randomised writing-task study, ChatGPT support raised essay-score gains, whereas knowledge gain and transfer showed no significant difference [19]. Performance gains thus co-exist with altered self-regulated learning and technology dependence. Molenaar frames hybrid human-AI learning technologies as a control-allocation problem, where AI systems detect, diagnose, and act, while levels of automation specify how control shifts between learner, teacher, and technology [20]. Assessment design has moved in the same direction. The AI Assessment Scale specifies graded levels of permitted GenAI use, from no AI to full AI, including a level where AI completes parts of a task and students evaluate the output [21]. Two-lane assessment guidance separates secured assurance of learning from open AI-integrated work [22], while Curtis argues that an all-or-none settlement misses the need for constrained middle-lane uses of GenAI [23]. Outside education, Dell'Acqua et al. show why task scope matters, as consultants using GPT-4 performed better on tasks inside the AI frontier yet worse on a task outside it [24]. Taken together, these studies narrow the contribution of the present paper. The paper does not introduce verification, task-specific AI permissions, or control allocation as isolated ideas. It connects them through learner-side markers and the interface layer, and it adds functional legibility as the missing link between system operation and student judgement.

3. A Relationship-Quality Typology

The typology rests on two states, a mediating interface layer, and three observable markers (Figure 1).



Figure 1. The relationship-quality typology, showing the target state, the failure state, and the interface layer as design surface, each with observable markers.

3.1 Capability-returning use: the target state

Following Licklider, we use symbiosis as a historical starting point for thinking about close human-computer coupling [4]. The educational target, however, differs from Licklider's tight-coupling ideal, since student-AI interaction should preserve and strengthen the student's ability to act without the system. We therefore define the target state as capability-returning use. Students who use an AI system to explain a concept, reconstruct the explanation in their own words, and check it against the course material end the interaction more capable than before. The defining criterion is the return, whereby the interaction produces an output and a gain on the human side. This criterion shifts the question from whether AI was used to whether the use returned capability to the student. The three markers introduced in Section 3.3 record the interaction pattern that plausibly precedes this return, and they do not measure the return itself. Section 6 treats this gap as the first task for empirical work.

3.2 Dependent use: the failure state

The failure state is the inverse, namely delegation without return. Risko and Gilbert describe the mechanism as cognitive offloading, in which externalising parts of a task can reduce the processing through which the corresponding skill is practised or maintained [5]. Applied to studying, a systematic review associates over-reliance on AI dialogue systems with weaker cognitive abilities [6], and survey evidence associates heavy AI-tool use with lower critical-thinking scores [7]. We call the resulting state dependence, since the student can no longer produce the delegated performance alone. Two properties make dependence a design problem. It remains silent, as the missing capability surfaces only when the system is unavailable, typically in an examination or a workplace. It plausibly builds up through many individually reasonable delegations, and the cross-sectional evidence cited above supports associations rather than trajectories, so the longitudinal designs named in Section 6 can test the temporal reading. Both properties call for continuous countermeasures at the interface and course level.

3.3 The interface layer and three observable markers

Between the two states sits the layer that shapes which of them forms, namely prompts, defaults, labels, checkpoints, and course rules. A frictionless copy-paste path makes wholesale delegation the default, while a visible verification step makes the capability-returning move the default. This is the paper's central design hypothesis. It suggests that relationship quality is co-produced by the student together with the surrounding interface and course design, which makes it a target for intervention and empirical testing. The same position underlies human-centred AI design [8] and hybrid-intelligence research [9], and the typology applies it to the concrete surfaces of AI-supported studying.

For design and for assessment, the typology needs markers that are observable without psychometric instruments. Three markers make the two states observable in interaction.

1. Verification. Do students check AI output against other sources or their own understanding before using it? Verification is the behavioural core of the critical-evaluation competence that AI-literacy frameworks demand [12], [14].

2. Task-scoped delegation. Is the AI tied to a defined role in a concrete task, such as summarise this text or critique this draft, or does it produce the performance wholesale? Scoped delegation preserves the return that defines capability-returning use, while wholesale delegation removes it.

3. Dependence awareness. Do students report being able to produce the performance without the system? The marker relies on self-assessment and serves as a risk indicator, since reported dependence signals that the delegated capability may no longer be available unaided. Self-assessments can misjudge actual capability, so Section 6 names calibration against demonstrated capability as a validation step.

Table 1 connects the three markers, together with two design conditions imported from transparency and governance work, to the design propositions of Section 5.

Marker or design condition	Type	Observable question	Design proposition
Verification	Marker	Do students check AI output before using it?	P1 verification at the point of output
Task-scoped delegation	Marker	Is the AI's role in the task defined and bounded?	P3 delegation contracts per assignment
Dependence awareness	Marker	Do students report being able to produce the work without AI?	P4 without-AI checkpoints and usage mirrors
Functional legibility	Design condition	Can students tell which component produced an output?	P2 labels and attribution at the point of input
Social anchoring	Design condition	Did the cohort commit to shared rules it helped write?	P5 co-created AI charters

Table 1. Relationship-quality markers, design conditions, and corresponding design propositions.

4. Descriptive Illustration from Two Open Datasets

The descriptive step asks whether the relationship states named by the typology appear in student survey data. We use the two openly licensed self-report datasets for a bounded descriptive purpose, namely to identify profile heterogeneity and to compare use patterns across item-based groups.

Dataset A is a single-institution survey on study habits and AI use, collected with proportional allocation across 14 undergraduate programs and quota-based fieldwork within programs (n = 357). The public package includes the raw Spanish file, a cleaned English analytical file, a codebook, and the questionnaire [10]. Dataset B is a global survey of higher-education students' perceptions of ChatGPT, collected between October 2024 and February 2025 (22,963 responses from 120 countries and territories, CC BY 4.0) [2]. For Dataset B, we removed 127 exact duplicate rows and 205 further rows that repeated the substantive item blocks of an earlier response. After this cleaning, 19,382 respondents provide a valid answer on general ChatGPT use, and of these 17,178 also provide complete ethical-concern and satisfaction scores. The comparison below assumes that this item-level missingness is unrelated to the pattern of interest. Both datasets rely on self-report, and Dataset A contains a large share of frequent AI users.

Dataset A supplies the profile check. Of 357 students, 218 (61 %) report using generative AI for study purposes almost always or always. Within this high-use band, we coded two markers at the two most-affirming answers on their published scales. Verification counts when students report checking AI output almost always or always; perceived dependence counts when they answer probably yes or a lot to the dependence item. Dataset A contains no item that captures task-scoped delegation, so the illustration covers two of the three markers. Cross-classifying two binary markers partitions the 218 students into four cells by construction, and the informative result is the cell sizes. 63 students (29 %) verify and report no dependence, the combination closest to capability-returning use in these data. 73 (33 %) verify and simultaneously report dependence, a combination in which checking coexists with a dependence signal and which the two-state typology does not predict. 38 (17 %) report dependence without verification, the combination closest to the failure state. 44 (20 %) report neither marker (Figure 2). All four cells are substantial, including the two combinations outside the named states. Usage frequency alone collapses these differences into one high-use category, while the two markers separate four distinct marker-response combinations.

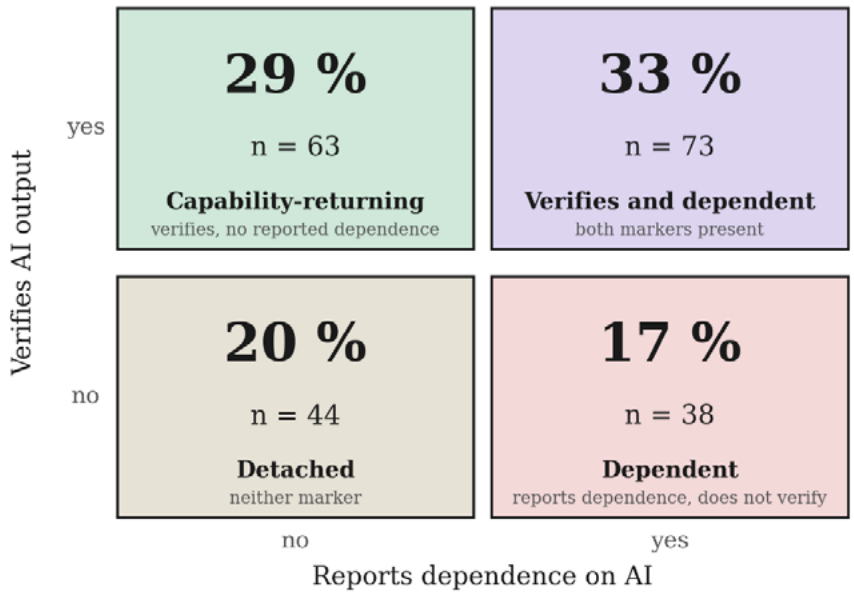


Figure 2. Four marker-response combinations among the 218 high-frequency users in Dataset A, cross-classified by the verification and dependence markers.

Dataset B adds the concern-use check. This comparison uses the 17,178 respondents with valid general-use, ethical-concern, and satisfaction scores after duplicate removal. We split respondents into thirds by their average satisfaction score and by their average ethical-concern score. Among highly satisfied students, the share of intensive users, defined as the two highest categories on the general-use item, is 48 % when ethical concern is low and 38 % when ethical concern is high. Within this subgroup, higher ethical concern accompanies somewhat lower intensive use, although intensive use remains common even under strong concern.

The same dataset locates everyday ChatGPT use in core academic tasks. The highest Q18 item means on the five-point frequency scale are summarising texts (3.0), brainstorming (3.0), academic writing (2.9), research assistance (2.9), and study assistance (2.8). These item means do not measure learning quality; instead, they show where design interventions would meet actual use, namely in recurring academic work rather than in peripheral experimentation.

Together, the two datasets support the descriptive premise of the typology. Dataset A shows that students with the same high usage frequency separate into different relationship profiles. Awareness of risks therefore co-exists with sustained use. These counts give the propositions an empirical starting point. Future work should develop the single-item markers into validated constructs and test the propositions in experimental, course-based, and longitudinal designs.

5. Design Propositions

The propositions make the three markers designable. They specify where an AI-supported course or interface must intervene if it aims to move student AI use from mere frequency toward relationship quality.

P1. Verification at the point of output. Interfaces should make checking the next action after generation, whether through source comparison, claim-basis inspection, or a required record of one verification step per AI-assisted task. AI-literacy frameworks require critical evaluation of AI output [12], [14], and this proposition locates that evaluation at the moment when students can still revise their judgement. Evidence from AI-assisted decision-making supports this timing, since cognitive forcing functions that interrupt default acceptance reduce overreliance more effectively than simple explanations, although they can reduce perceived usability [18]. For learning environments, verification therefore needs to appear at the point of reuse, not only as a general competence outcome.

P2. Functional legibility. Learning environments that combine several AI functions should identify, at the point of input and output, which component processes the input, what operation it performs, and what kind of output it returns. Students can evaluate an output only if the interface makes its operation legible, whether it transcribed, retrieved, generated, summarised, or recommended the answer. This links the proposition directly to interpretive affordances in AI literacy [12], to UNESCO's requirements for transparent, explainable, human-controlled AI use [14], and to Shneiderman's account of comprehensible, predictable, and controllable automation [8].

P3. Task-scoped delegation contracts. Assignments should define the AI role per task, meaning which steps may be delegated, which steps must remain the students' own, and how the work returns to the student. Assessment-scale work already defines graded permission levels for GenAI use, including levels where AI supports ideation, editing, task completion, or full task work [21]. The delegation contract proposed here uses that permission logic for a narrower

design purpose, since it ties each permitted AI role to the learner-side return that must follow. For example, the AI may draft counter-arguments, whereas students must select, revise, and defend the final position. This gives assessment a concrete object, namely the declared division of labour and the students' documented hand-back, rather than an unverifiable suspicion about hidden AI use.

P4. Dependence-awareness feedback. Courses and systems should mirror delegation back to the learner through periodic without-AI checkpoints or usage summaries that show which task types students have not completed unassisted. Two-lane assessment guidance protects assurance of learning through secured AI-free tasks while allowing AI-integrated work elsewhere [22]. Curtis's critique of all-or-none assessment shows why many tasks require constrained AI use rather than total prohibition or unrestricted use [23]. Without-AI checkpoints translate that assessment logic into feedback for the learner, since they reveal which performances still work without the system while there is time to adjust practice. This proposition also follows from human-centred AI competence, which treats human agency and avoidance of AI dependence as part of responsible use [14], and from Shneiderman's warning that automation can undermine skills needed when automation fails [8].

P5. Co-created AI-use rules. Institutions should combine clear institutional boundaries with cohort-level AI-use charters. Students and teachers should specify permitted AI roles task by task, record the resulting rules, and revisit them when assignments change. As Section 4 shows, awareness of risks co-exists with sustained use, and a charter addresses this co-existence by making students practise the judgement that AI competence requires [14] and the self-organised action that Ehlers places at the centre of future capability in emergent contexts [15].

6. Limitations and Future Work

The argument rests on a conceptual typology and a descriptive illustration. Licklider grounds the symbiosis vocabulary in close human-computer coupling, where humans set goals, define criteria, and evaluate results while computers take over routinisable work [4]. This paper transfers that vocabulary to student-AI relationships and uses it as design terminology.

The empirical illustration reports counts, shares, and item means from self-report surveys. Its evidential role is to make relationship combinations visible, not to estimate effects. The thresholds for the profile markers use the two most-affirming response options on each published scale, and independent studies should test whether these settings remain stable. Dataset A covers one institution and contains many frequent users. Dataset B captures ChatGPT use and therefore represents one prominent system rather than the full range of generative AI tools. The markers also remain single survey items and require validation as constructs. Above all, the markers capture the interaction pattern, not the capability return that defines the target state. Whether bounded, checked delegation predicts measured capability gains requires pre and post capability tests, and this comparison is the first task for future work.

Future work follows directly from these boundaries. Verification, task-scoped delegation, and dependence awareness should be developed into short validated measures and compared with behavioural traces. Interface studies can test verification affordances, functional labels, and delegation feedback as controlled variants. Course-level studies should examine delegation contracts and co-created AI-use rules inside real assessment practices, where these propositions gain practical meaning. Longitudinal designs can then test whether early relationship combinations predict later competence development.

7. Conclusion

GenAI use in higher education is usually discussed through quantity, meaning how many students use it, how often, and for which tasks. This paper shifted the unit of analysis to relationship quality. The proposed typology distinguishes forms of use in which AI returns capability to the student from forms in which students report that they can no longer produce the delegated performance alone. It also locates interface and course design as part of the relationship rather than as neutral surroundings. The descriptive survey illustration gives the typology empirical traction, since students with similar usage intensity fall into different relationship combinations, and ethical concern can co-exist with intensive use. The five propositions translate this argument into design decisions for AI-supported learning environments. The resulting task is concrete, namely to design student-AI interaction so that repeated use strengthens the student's own capacity to judge, act, and learn.

8. Acknowledgements

This research was supported by the EUonAIR project, no. 101177370, ERASMUS-EDU-2024-EUR-UNIV-1, within the framework of the EUonAIR Centre of Excellence in Responsible AI in Education (CORE.AI), co-funded by the European Union.

9. Ethics and data availability

This paper reports a secondary analysis of two openly licensed, anonymised survey datasets [2], [10], and the authors collected no new data from human participants. Both datasets are available from Mendeley Data, doi 10.17632/mcwb2ppdsw.1 (Dataset A) and doi 10.17632/nv2343nwsb.1 (Dataset B, CC BY 4.0). A Python script that reproduces every reported figure is available from the authors upon request.

References

[1] D. Ravšelj et al., "Higher education students' perceptions of ChatGPT: A global study of early reactions", *PLOS ONE*, Volume 20, Issue 2, e0315011, 2025.

[2] A. Aristovnik et al., "Higher Education Students' Evolving Perceptions of ChatGPT: Global Survey Data from the Academic Year 2024-2025", *Mendeley Data*, V1, 2025.

[3] M. Bower, J. Torrington, J. W. M. Lai, P. Petocz and M. Alfano, "How should we change teaching and assessment in response to increasingly powerful generative Artificial Intelligence? Outcomes of the ChatGPT teacher survey", *Education and Information Technologies*, Volume 29, pg. 15403-15439, 2024.

[4] J. C. R. Licklider, "Man-Computer Symbiosis", *IRE Transactions on Human Factors in Electronics*, Volume HFE-1, Issue 1, pg. 4-11, 1960.

[5] E. F. Risko and S. J. Gilbert, "Cognitive Offloading", *Trends in Cognitive Sciences*, Volume 20, Issue 9, pg. 676-688, 2016.

[6] C. Zhai, S. Wibowo and L. D. Li, "The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review", *Smart Learning Environments*, Volume 11, Article 28, 2024.

[7] M. Gerlich, "AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking", *Societies*, Volume 15, Issue 1, Article 6, 2025.

[8] B. Shneiderman, "Human-Centered Artificial Intelligence: Reliable, Safe and Trustworthy", *International Journal of Human-Computer Interaction*, Volume 36, Issue 6, pg. 495-504, 2020.

[9] D. Dellermann, P. Ebel, M. Söllner and J. M. Leimeister, "Hybrid Intelligence", *Business & Information Systems Engineering*, Volume 61, pg. 637-643, 2019.

[10] C. D. Correa Alvarez, "Study habits and artificial intelligence use among university students: A proportionally stratified survey dataset", *Mendeley Data*, V1, 2026.

[11] W. Holmes, M. Bialik and C. Fadel, "Artificial Intelligence in Education", *Globethics Publications*, excerpt, orig. Center for Curriculum Redesign, Boston, 2019, 2023.

[12] D. Long and B. Magerko, "What is AI literacy? Competencies and design considerations", *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pg. 1-16, 2020.

[13] D. T. K. Ng, J. K. L. Leung, S. K. W. Chu and M. S. Qiao, "Conceptualizing AI literacy: An exploratory review", *Computers and Education: Artificial Intelligence*, Volume 2, Article 100041, 2021.

[14] UNESCO, *AI competency framework for students*, UNESCO Publishing, Paris, 2024.

[15] U.-D. Ehlers, *Future Skills: Lernen der Zukunft - Hochschule der Zukunft*, Springer VS, Wiesbaden, 2020.

[16] J. Kim, H. Lee and Y. H. Cho, "Learning design to support student-AI collaboration: perspectives of leading teachers for AI in education", *Education and Information Technologies*, Volume 27, pg. 6069-6104, 2022.

[17] A. Nguyen, Y. Hong, B. Dang and X. Huang, "Human-AI collaboration patterns in AI-assisted academic writing", *Studies in Higher Education*, Volume 49, Issue 5, pg. 847-864, 2024.

[18] Z. Buçinca, M. B. Malaya and K. Z. Gajos, "To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-Assisted Decision-Making", *Proceedings of the ACM on Human-Computer Interaction*, Volume 5, Issue CSCW1, Article 188, 2021.

[19] Y. Fan, L. Tang, H. Le, K. Shen, S. Tan, Y. Zhao, Y. Shen, X. Li and D. Gašević, "Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance", *British Journal of Educational Technology*, Volume 56, Issue 2, pg. 489-530, 2025.

[20] I. Molenaar, "Towards hybrid human-AI learning technologies", *European Journal of Education*, Volume 57, Issue 4, pg. 632-645, 2022.

[21] M. Perkins, L. Furze, J. Roe and J. MacVaugh, "The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment", *Journal of University Teaching & Learning Practice*, Volume 21, Issue 6, 2024.

[22] D. Liu and A. Bridgeman, "What to do about assessments if we can't out-design or out-run AI?", *Teaching@Sydney*, 12 July 2023, available at: https://educational-innovation.sydney.edu.au/teaching@sydney/what-to-do-about-assessments-if-we-cant-out-design-or-out-run-ai/

[23] G. J. Curtis, "The two-lane road to hell is paved with good intentions: why an all-or-none approach to generative AI, integrity, and assessment is insupportable", *Higher Education Research & Development*, Volume 44, Issue 8, pg. 2151-2158, 2025.

[24] F. Dell'Acqua et al., "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of Artificial Intelligence on Knowledge Worker Productivity and Quality", *Organization Science*, Articles in Advance, 2026.

THE USE OF AI TOOLS

The authors used ChatGPT (OpenAI) and Claude (Anthropic) for language editing and translation of selected parts of the manuscript. All content was reviewed and verified by the authors.