

Why University Students Choose Generative AI for Academic Tasks: An Exploratory Mixed-Methods Factor Analysis

Astried Nurrizqy Savanah^{1*}, Mohammad Ilham Triananda², Rifky Ramadhandi Anandithya³, Samuel Christian⁴, Muhammad Nur Azam Ghazali⁵

^{1,2,3,4}Sekolah Tinggi Manajemen PPM Jakarta, Indonesia

⁵Universiti Kebangsaan Malaysia, Malaysia

*Corresponding author, e-mail: astried.nurrizqy99@gmail.com

Abstract

This exploratory sequential mixed-methods study examines how university students evaluate generative artificial intelligence when choosing tools for academic tasks. Qualitative interview coding informed a 24-item questionnaire, followed by exploratory factor analysis of 70 undergraduate and master's students. Seventy-two responses were collected and 70 were included in the factor analysis; the reason for excluding two responses was not recorded. Sampling adequacy was modest (KMO = .560), Bartlett's test was significant, and Cronbach's alpha was .668 (.695 standardized). Seven components with eigenvalues above one explained 65.097% of observed variance: reliability, multifunctionality, exploratory capacity, suitability, prompt understanding, information clarity, and awareness. The findings show that student choice combines appraisal of the tool, fit with the academic task, and self-appraisal of the user's ability to prompt, interpret, verify, and use generated material responsibly. The study therefore reframes adoption as evaluative AI literacy rather than simple convenience. Because the sample is small and diagnostic values are modest, the factor structure should be replicated with larger samples and confirmatory analysis.

Keywords: academic tasks; AI literacy; generative artificial intelligence; university student.

How to Cite: Savanah, A. N. et al. (2024). Why University Students Choose Generative AI for Academic Tasks: An Exploratory Mixed-Methods Factor Analysis. *Charta Educa: Jurnal Kajian Pendidikan*, 1(2), 46-54.



This is an open access article distributed under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2024 by author.

Introduction

Generative artificial intelligence has moved rapidly from a specialist technology to an everyday academic tool. University students now use conversational systems to brainstorm topics, summarize readings, translate concepts, generate examples, revise prose, and troubleshoot code. This expansion makes adoption an educational question rather than a purely technical one. The important issue is not simply whether students use generative AI, but what judgments they make when deciding which tool or output is acceptable for an academic task. Research published during the early diffusion of ChatGPT consistently identified a tension between usefulness and uncertainty: generative systems can support explanation, ideation, and feedback, yet they can also produce inaccurate information, fabricated references, overconfident language, and outputs that encourage uncritical dependence (Baidoo-Anu & Ansah, 2023; Kasneci et al., 2023; Lo, 2023).

Student choice is therefore better understood as an evaluative process. Convenience matters, but so do perceived reliability, task fit, transparency, communicative clarity, and the learner's confidence in formulating prompts and checking answers. Studies of university students have shown that perceived benefits coexist with concerns about accuracy, privacy, academic integrity, and the erosion of independent work (Chan & Hu, 2023; Cotton et al., 2024). These concerns complicate simple acceptance models in which a useful and easy-to-use

technology is assumed to be adopted. With generative AI, the user must continually judge the quality of a probabilistic response and decide what requires verification.

This evaluative dimension connects technology adoption with AI literacy. AI literacy in higher education includes understanding what generative systems can and cannot do, designing prompts that elicit relevant responses, comparing generated claims with reliable sources, recognizing hallucination and bias, and taking responsibility for the final academic product. Prompting is therefore not merely a technical interaction skill. It is part of a broader epistemic practice in which students frame a problem, inspect evidence, and decide whether a generated response is trustworthy enough to use. Policy frameworks for university AI use increasingly emphasize this combination of practical competence, critical verification, and ethical judgment (Chan, 2023; Perkins, 2023).

The study addresses this problem empirically by moving from qualitative interviews to a questionnaire and exploratory factor analysis. Rather than imposing a predetermined adoption model, interview coding was used to develop 24 survey items, allowing students' own considerations to inform the quantitative stage. The factor analysis then examined whether those considerations clustered into interpretable dimensions. This design is useful because it treats student decision making as multidimensional, while its modest sample means the resulting structure must remain exploratory rather than definitive.

The study asks two questions: What dimensions organize students' considerations when they choose generative AI for academic tasks, and what do those dimensions imply for AI literacy in higher education? Its contribution is to show that adoption is not reducible to efficiency. Reliability, multifunctionality, exploratory capacity, suitability, prompt understanding, information clarity, and awareness jointly indicate that students choose AI through a combination of tool appraisal and self-appraisal. The educational implication is that universities need to teach students not only how to access AI, but how to interrogate it.

Method

This study used an exploratory sequential mixed-methods design to identify the considerations that university students use when choosing generative artificial intelligence for academic work. The first phase was qualitative: in-depth interviews were coded to identify recurring considerations in students' descriptions of AI use. Those codes were then translated into questionnaire statements. The second phase was quantitative: a 24-item Likert questionnaire was administered to students with experience using generative AI for academic tasks, and the resulting item responses were examined through exploratory factor analysis in SPSS. The design was therefore developmental rather than experimental; the interviews informed instrument construction, and the survey was used to explore the latent organization of the considerations identified in the first phase.

Seventy-two questionnaires were collected, of which 70 were included in the factor analysis. The exclusion reason for two responses was not recorded. The analyzed sample consisted of 48 undergraduate and 22 master's students. This sample is adequate for describing the respondents and for a preliminary exploration of item clustering, but it is small relative to common recommendations for stable factor recovery. The analysis is therefore treated as exploratory and hypothesis-generating rather than as validation of a definitive measurement model.

Factorability checks yielded a Kaiser–Meyer–Olkin value of .560 and a statistically significant Bartlett's test of sphericity. Internal consistency for the 24-item scale was Cronbach's $\alpha = .668$, with a standardized alpha of .695. These values indicate that the items shared sufficient covariance to justify exploratory analysis while also signaling that the scale should not be interpreted as a highly homogeneous instrument. Components with eigenvalues greater than one were retained; seven components together accounted for 65.097% of the observed variance. Because alternative extraction and rotation comparisons were not reported, the seven-component solution is interpreted cautiously.

Interpretation proceeded in two stages. First, the seven component labels reported in the study—reliability, multifunctionality, exploratory capacity, suitability, prompt understanding, information clarity, and awareness—were treated as descriptive summaries of the item clusters rather than as fixed psychological traits. Second, the labels were compared with higher-education research on generative AI to examine whether they reflected broader issues of usefulness, verification, interaction quality, academic integrity, and AI literacy. This separation between statistical output and educational interpretation is important because a factor label is an analytic interpretation, not a directly observed property of a student.

A formal institutional ethics approval number and the precise consent procedure were not reported. This is a methodological reporting limitation. Future confirmatory work with the dataset should document recruitment, consent, response screening, missing-data handling, and the exact factor-extraction and rotation settings.

Results and Discussion

Sampling adequacy supported exploration, not scale validation

The KMO value of .560 and the significant Bartlett's test indicate that the correlation matrix contained enough shared variance to justify exploratory factor analysis. At the same time, .560 is a modest KMO value. The statistical evidence therefore supports a preliminary search for structure but not a claim that the seven-factor solution is stable across populations. Cronbach's alpha of .668 (.695 standardized) points in the same direction: the 24 items were sufficiently related to be considered together, yet the scale did not behave as a single strongly homogeneous construct. This pattern is consistent with the study's premise that students evaluate AI through several distinct considerations rather than through one global attitude.

Seven components explained 65.1% of the observed variance

The retained components together explained 65.097% of variance and were interpreted as reliability, multifunctionality, exploratory capacity, suitability, prompt understanding, information clarity, and awareness. The seven labels are educationally meaningful because they span three different objects of judgment. Reliability and information clarity concern the perceived quality of the output; multifunctionality, exploratory capacity, and suitability concern the relationship between the tool and the academic task; prompt understanding and awareness concern the learner's own capacity to interact with the system and recognize its limits.

Students' choices combined tool appraisal and self-appraisal

The factor pattern indicates that students did not choose generative AI solely because it was fast or convenient. They also attended to whether the system appeared dependable, whether its response was understandable, whether it could support several forms of academic work, and whether it suited the problem at hand. Equally important, prompt understanding and awareness placed part of the decision inside the user rather than the technology. Students' choices thus involved self-appraisal: they had to judge whether they could formulate an adequate request, interpret the response, and recognize when additional verification was necessary.

The factors describe an evaluative use process

Taken together, the seven components can be read as stages in an evaluative use process. A student identifies a task, considers whether an AI system is suitable and sufficiently multifunctional, formulates a prompt, examines the clarity and reliability of the answer, uses the system to explore alternatives, and remains aware that generated material may require verification or revision. The data do not establish that every respondent followed this sequence consciously, but the structure helps explain why adoption cannot be reduced to access. The educational issue is the quality of judgment exercised during use.

Student adoption research published before the end of 2024 reinforces the importance of treating use as a judgment process rather than a binary choice. Surveys and acceptance studies show that students' engagement with generative AI varies with perceived usefulness, facilitating conditions, familiarity, task fit, and concerns about responsible use (Balabdaoui et al., 2024; Baidoo-Anu et al., 2024; Strzelecki, 2024; Yu et al., 2024). Higher-education commentary also emphasizes that benefits depend on how learners evaluate opportunities and limitations rather than on access alone (Fuchs, 2023). These patterns support interpreting the seven components as dimensions of evaluative AI literacy.

Discussion

From adoption to evaluative AI literacy

The main contribution of the study is to shift attention from whether students adopt generative AI to how they evaluate it while using it. Early higher-education research showed that students rapidly recognized the convenience and flexibility of conversational AI, but also expressed uncertainty about accuracy, ethics, and the appropriate role of AI in assessed work (Chan & Hu, 2023; Firat, 2023; Lo, 2023). The seven factors found here give that tension a more differentiated structure. Students' decisions combine judgments about the system, the task, and their own interaction competence. In that sense, AI adoption is already a literacy problem: the learner must know enough about the tool to decide when its output can be used, when it should be questioned, and when another source or method is preferable.

This interpretation is consistent with broader accounts of AI literacy in higher education. Generative AI differs from many earlier educational technologies because its output is probabilistic, conversational, and rhetorically fluent. Fluency can create an impression of authority even when factual support is weak. The educational response therefore cannot be limited to technical access or software training. Students need habits of inquiry that make verification routine: identifying the type of claim being generated, checking important claims against authoritative sources, comparing alternative explanations, tracing citations, and distinguishing a plausible formulation from a defensible academic argument (Kasneci et al., 2023; Tlili et al., 2023).

Reliability is relational rather than purely technical

Reliability emerged as a central consideration, but it should not be understood only as a fixed property of the AI model. In academic work, reliability is relational: it depends on the question asked, the domain, the availability of external evidence, and the user's verification practices. A response that is adequate for brainstorming possible keywords may be inadequate as evidence for a literature review. A generated explanation may help a student recognize a concept but still require confirmation in a textbook or peer-reviewed source. This distinction matters because students who treat reliability as an all-or-nothing characteristic may either trust the system too much or reject useful forms of assistance too quickly.

Research on large language models has repeatedly emphasized this double character. Generative systems can support explanation, idea generation, language development, and formative feedback while also producing hallucinations, biased framings, and fabricated references (Baidoo-Anu & Ansah, 2023; Farrokhnia et al., 2024; Kasneci et al., 2023). The practical implication is to teach calibrated trust. Students should learn to match the intensity of verification to the stakes of the task. Low-stakes ideation may tolerate exploratory output; claims used in assessment, scientific reasoning, or professional decision-making require stronger source checking.

Multifunctionality and suitability should be considered together

Multifunctionality reflects the appeal of a tool that can summarize, translate, explain, generate examples, reorganize prose, and support coding or problem solving. Yet the existence of many functions does not mean that all functions are equally appropriate in every academic context. Suitability therefore moderates multifunctionality. Students need to ask not only "Can the system do this?" but "Should this task be delegated, supported, or retained as my own cognitive work?" That question becomes particularly important when the task itself is intended to develop the skill that AI could automate.

For example, using AI to generate alternative examples after a student has attempted an explanation may extend learning, whereas asking the system to produce the entire explanation before the student engages with the concept may reduce productive struggle. Similarly, language support can widen participation for students writing in an additional language, but uncritical rewriting can obscure whether the final text represents the student's own reasoning. These tensions explain why academic AI policy increasingly distinguishes permitted support from substitution of assessed intellectual work (Chan, 2023; Perkins, 2023).

Exploratory capacity can support inquiry when it remains provisional

The exploratory-capacity factor is pedagogically promising because conversational AI can rapidly generate alternative perspectives, examples, counterarguments, or search terms. Used well, this capacity can lower the cost of beginning an inquiry. A student who is unsure how to frame a topic can test several formulations; a student who has encountered a difficult concept can request examples at different levels of complexity; a student preparing a discussion can generate counterpositions to evaluate. The value lies in expansion of the inquiry space rather than automatic production of the final answer.

The risk appears when exploration becomes epistemic closure. Because a conversational system responds immediately, students may stop searching when they receive the first coherent answer. This can narrow rather than widen inquiry. Effective use therefore requires a deliberate pause between generation and acceptance: asking what is missing, requesting alternative accounts, comparing AI output with course literature, and documenting which parts of the final argument are supported by traceable sources. These practices align with calls to treat human-AI collaboration as a form of augmented work in which human judgment remains decisive (Grønsund & Aanestad, 2020; Dwivedi et al., 2023).

Prompt understanding is an academic interaction skill

Prompt understanding was not merely a convenience variable. It represents the learner's ability to translate an academic need into an interaction that the system can interpret. Effective prompting requires specifying context, purpose, constraints, desired format, and sometimes the reasoning process expected from the model. Yet prompt skill should not be fetishized. A sophisticated prompt can still generate an inaccurate answer, and a technically polished interaction does not guarantee sound academic judgment. Prompt competence is therefore best taught together with evaluation competence.

This has implications for curriculum design. Rather than teaching lists of "perfect prompts," instructors can design activities in which students compare how changes in a prompt affect the response, identify unsupported claims, and revise both the prompt and the resulting answer. Such tasks make the interaction itself an object of reflection. They also reveal that prompting is recursive: the user interprets a response, recognizes a gap, reformulates the question, and evaluates the next output. This recursive process resembles inquiry more than simple information retrieval and can become educationally valuable when students are required to explain the basis for accepting or rejecting generated material.

Information clarity is useful but can conceal uncertainty

Students valued information clarity, which is unsurprising because conversational AI often produces well-structured prose. Clarity can support comprehension, especially when students ask for definitions, analogies, or stepwise explanations. However, rhetorical clarity is not equivalent to evidentiary quality. Generative systems may express uncertain or incorrect claims in confident language. Higher education should therefore help students separate readability from validity. A clear answer should be easier to interrogate, not easier to accept uncritically.

One practical strategy is to require evidence mapping. Students can highlight factual claims in an AI response and identify which ones can be verified through scholarly or authoritative sources. Another strategy is comparative prompting: students ask the system for multiple explanations and then evaluate differences using disciplinary criteria. These activities reposition clarity as a starting point for analysis. They also address a concern identified across the early ChatGPT literature: the technology can make weak information appear academically polished, creating new demands for critical reading and source evaluation (Cotton et al., 2024; AlAfnan et al., 2023).

Awareness connects AI literacy with academic integrity

The awareness factor links practical use to responsibility. Students need to know that AI output can contain error, bias, unattributed synthesis, or material that conflicts with course rules. They also need to recognize when AI assistance should be disclosed. Academic integrity in the generative-AI era is therefore not adequately addressed by a binary prohibition model. The more sustainable approach is to specify what forms of assistance are permitted for particular tasks and require students to remain accountable for the accuracy and originality of submitted work (Perkins, 2023; Cotton et al., 2024).

Institutional policy should be paired with assessment design. If an assignment rewards only a polished final product, students have little incentive to make their reasoning process visible. Assessments can instead include drafts, annotated sources, oral explanation, process logs, or reflective accounts of how tools were used. These design choices do not eliminate inappropriate AI use, but they align assessment with learning and make student judgment more observable. Policy, pedagogy, and assessment therefore need to develop together rather than treating AI use as a standalone disciplinary problem (Chan, 2023; Dwivedi et al., 2023).

Interpreting the seven-factor solution cautiously

The factor solution should not be over-generalized. The sample of 70 is small for a 24-item exploratory analysis, the KMO value is modest, and the internal-consistency estimate is moderate. Seven factors explaining 65.1% of variance may represent a useful descriptive structure, but small samples can produce unstable loading patterns. The result is best viewed as a framework for further measurement development. A larger study should test whether the seven dimensions recur, whether some factors merge, and whether items behave similarly across undergraduate and master's students.

Future work should report extraction and rotation choices in full, inspect communalities and cross-loadings, and use a new sample for confirmatory factor analysis rather than validating the model on the same respondents. It would also be useful to connect factor scores with observed behavior. Students may say that reliability matters while still accepting unverified answers under time pressure. Combining survey measures with task-based observations, prompt histories, or think-aloud protocols could show how evaluative criteria operate during actual AI-assisted work. Such evidence would move the field from perceptions of AI literacy toward enacted AI literacy.

Implications for universities

The seven factors offer a practical organizing frame for university support. Reliability and information clarity point to verification and critical-reading skills; multifunctionality and exploratory capacity point to purposeful task selection; suitability points to assessment and disciplinary fit; prompt understanding points to interaction design; and awareness points to ethics, disclosure, and accountability. These dimensions can be integrated into existing information-literacy, academic-writing, and research-methods teaching rather than isolated in a one-off technology workshop.

The study also suggests that AI literacy should remain disciplinary. What counts as adequate verification differs across fields. A history student may need to trace primary and secondary sources; a management student may need to assess the evidentiary basis of a recommendation; a quantitative student may need to reproduce calculations or code; a language student may need to distinguish editing support from authorship. General AI literacy provides a common foundation, but responsible use becomes meaningful only when connected to the standards of evidence and reasoning that define a field (Zawacki-Richter et al., 2019; González-Calatayud et al., 2021).

Instrument development is part of the contribution

The qualitative-to-quantitative sequence is itself important because it begins with students' language rather than assuming that established technology-acceptance constructs fully describe generative AI. Traditional

acceptance models remain useful for questions of usefulness and ease, but generative systems introduce additional concerns about hallucination, authorship, prompt formulation, and the relationship between fluent output and reliable evidence. Developing questionnaire items from interviews therefore gave the instrument a closer connection to the emerging practices students were actually negotiating. At the same time, future instrument development should publish the interview coding process and item-to-code mapping so readers can evaluate content validity.

A stronger next phase would combine expert review, cognitive interviewing, and pilot testing before large-scale factor analysis. Expert review can examine whether items adequately represent the intended dimensions; cognitive interviewing can identify how students interpret ambiguous wording; pilot data can reveal floor or ceiling effects. Such steps are particularly valuable in a rapidly changing domain because the meaning of “AI use” can shift as tools and institutional rules change. Measurement development should therefore be iterative rather than assuming that a factor structure obtained during early adoption will remain fixed.

Undergraduate and master’s students may use AI differently

The analyzed sample included 48 undergraduate and 22 master’s students, but the study was not large enough to compare these groups reliably. That limitation points to a useful research question. More experienced students may have different academic tasks, disciplinary knowledge, source-evaluation habits, or tolerance for uncertainty. They may use AI more selectively for synthesis and critique, whereas less experienced students may value explanation and task completion. These are hypotheses, not findings of the present study, but they illustrate why future validation should examine measurement invariance and group differences rather than treating university students as a homogeneous user population.

Disciplinary variation may be equally important. Reliability in a quantitative course may involve reproducible calculations; in a history course it may involve traceable sources and contextual accuracy; in language learning it may involve nuance, register, and communicative purpose. Suitability and awareness are therefore likely to be shaped by disciplinary epistemology. University-wide policy can define common principles, but AI literacy instruction should be translated into the standards by which different academic communities judge evidence.

Human-in-the-loop use provides a practical model

The factor pattern can be summarized through a human-in-the-loop model of academic AI use. The system generates possibilities, but the student retains responsibility for defining the task, providing relevant context, evaluating the response, checking evidence, and deciding what enters the final work. Human-in-the-loop research emphasizes that augmentation changes the distribution of work between people and algorithms rather than eliminating human judgment (Grønsund & Aanestad, 2020). In education, that distribution should be designed so that AI supports learning without outsourcing the reasoning that an assessment is intended to develop.

This model also clarifies why bans and unrestricted permission are both blunt instruments. Some tasks can legitimately use AI for brainstorming or feedback, while others need protected spaces in which students demonstrate unaided competence. The educational challenge is to specify which cognitive work must remain visible and which forms of assistance can extend it. An assignment-by-assignment approach can therefore be more defensible than a single rule applied to all learning activities.

Policy should make verification and disclosure actionable

Institutional guidance is most useful when students can translate it into decisions. Statements such as “use AI responsibly” are too broad unless accompanied by examples of permitted support, unacceptable substitution, verification expectations, and disclosure formats. Chan’s (2023) policy framework and Perkins’ (2023) integrity analysis both point toward the need for explicit educational guidance rather than policy language alone. The seven factors can help structure such guidance by asking students to document task suitability, the purpose of AI use, how outputs were checked, and what human contribution shaped the final submission.

Universities can also integrate AI literacy into research and information-literacy programs. Librarians, writing centers, disciplinary lecturers, and academic-integrity units each address different parts of the evaluative process. Coordinated provision would help students see that source verification, writing decisions, tool use, and academic responsibility are connected. Such coordination is consistent with the broader observation that successful educational AI adoption requires educators and institutional systems, not only technologies (Zawacki-Richter et al., 2019).

Limits define the next measurement agenda

Several limitations should guide interpretation. The sample was small, recruitment details are incompletely documented, two collected responses were excluded without an available rationale, and the KMO and reliability values were modest. The seven-factor solution may therefore contain sample-specific structure. The study also measured stated considerations rather than directly observing how students behave during an authentic

assignment. These limits do not invalidate the exploratory result, but they prevent the factors from being treated as a validated scale.

A robust next study would recruit a larger, diverse sample; pre-register the analytic plan; report the full loading matrix, communalities, rotation, and item-retention criteria; and confirm the structure on an independent sample. Behavioral tasks could then test whether students who score highly on reliability or awareness actually verify outputs more effectively. Connecting self-report with observed performance would be a major step toward a defensible measure of evaluative AI literacy.

A practical student workflow for responsible use

The findings can be translated into a simple student workflow: define the academic purpose, decide whether AI assistance is suitable, construct a prompt that includes relevant context, treat the generated response as provisional, check important claims against authoritative sources, revise the output using disciplinary criteria, and disclose assistance when course rules require it. Each step corresponds to one or more of the identified factors. Suitability governs whether the tool should be used; prompt understanding governs interaction; reliability and information clarity govern evaluation; exploratory capacity and multifunctionality govern what the tool contributes; awareness governs verification and responsibility. This workflow makes the factor solution actionable without presenting it as a validated diagnostic scale.

Teaching the workflow can also create a common language between students and instructors. Instead of debating AI use only in terms of prohibition or permission, teachers can ask students to explain why a tool was suitable, how an output was checked, and which intellectual decisions remained their own. These questions make evaluative judgment visible and align AI use with the purposes of higher education: learning to reason with evidence, communicate responsibly, and exercise informed independence.

Conclusion

This exploratory mixed-methods study identified seven dimensions that organized students' considerations when choosing generative AI for academic tasks: reliability, multifunctionality, exploratory capacity, suitability, prompt understanding, information clarity, and awareness. Together, the factors accounted for 65.1% of observed variance. Their importance lies less in establishing a final measurement model than in showing that student adoption combines tool appraisal with self-appraisal. Students evaluate what an AI system can do, whether it fits the task, how clearly and reliably it responds, and whether they themselves can prompt, interpret, verify, and use the output responsibly.

The findings support an educational shift from access-oriented AI training toward evaluative AI literacy. Universities should teach students to calibrate trust, verify consequential claims, select appropriate forms of assistance, disclose use where required, and preserve responsibility for the final academic product. Because the analyzed sample was small and the KMO and reliability values were modest, the seven-factor solution requires replication with larger samples and confirmatory analysis. Its strongest contribution is therefore conceptual and exploratory: generative-AI use in higher education is not simply a decision to adopt a tool; it is a sequence of judgments about evidence, task fit, interaction quality, and academic responsibility.

Declarations

Author contribution

Author 1: developed the main research idea and objectives, designed the research methodology, conducted data collection and investigation, performed the data analysis, and prepared the first draft of the manuscript. Author 2: contributed to the development and implementation of the research methodology, supported the data collection and investigation process, and reviewed and revised the manuscript to improve its academic quality, clarity, and structure. Author 3: provided input into the research concept and direction. Author 4: supervised and guided the overall research process. Author 5: reviewed and improved the manuscript, and contributed to obtaining financial support or research funding

Funding

The research received no specific grant from any funding agency.

Conflict of Interest

The authors declare that there is no conflict of interest related to the publication of this article.

Use of generative AI

Generative AI were used solely for translation and language editing of the manuscript. The authors take full responsibility for the content and accuracy of the article.

Acknowledgments

Authors would like to express their sincere gratitude to all parties who contributed to and supported the completion of this study. Their valuable assistance, support, and encouragement are greatly appreciated.

References

- AlAfnan, M. A., Dishari, S., Jovic, M., & Lomidze, K. (2023). ChatGPT as an educational tool: Opportunities, challenges, and recommendations for communication, business writing, and composition courses. *Journal of Artificial Intelligence and Technology*, 3(2), 60–68.
- Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, 7(1), 52–62.
- Baidoo-Anu, D., Asamoah, D., Amoako, I., & Mahama, I. (2024). Exploring student perspectives on generative artificial intelligence in higher education learning. *Discover Education*, 3, 98. <https://doi.org/10.1007/s44217-024-00173-z>
- Balabdaoui, F., Dittmann-Domenichini, N., Grosse, H., Schlienger, C., & Kortemeyer, G. (2024). A survey on students' use of AI at a technical university. *Discover Education*, 3, 51.
- Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education*, 20, 38. <https://doi.org/10.1186/s41239-023-00408-3>
- Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20, 43. <https://doi.org/10.1186/s41239-023-00411-8>
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228–239.
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., ... Wright, R. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
- Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*, 61(3), 460–474. <https://doi.org/10.1080/14703297.2023.2195846>
- Firat, M. (2023). What ChatGPT means for universities: Perceptions of scholars and students. *Journal of Applied Learning and Teaching*, 6(1), 57–63.
- Fuchs, K. (2023). Exploring the opportunities and challenges of NLP models in higher education: Is ChatGPT a blessing or a curse? *Frontiers in Education*, 8, 1166682.
- González-Calatayud, V., Prendes-Espinosa, P., & Roig-Vila, R. (2021). Artificial intelligence for student assessment: A systematic review. *Applied Sciences*, 11(12), 5467. <https://doi.org/10.3390/app11125467>
- Grønsund, T., & Aanestad, M. (2020). Augmenting the algorithm: Emerging human-in-the-loop work configurations. *The Journal of Strategic Information Systems*, 29(2), 101614.
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, 13(4), 410. <https://doi.org/10.3390/educsci13040410>
- Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching & Learning Practice*, 20(2), Article 7. <https://doi.org/10.53761/1.20.02.07>
- Strzelecki, A. (2024). Students' acceptance of ChatGPT in higher education: An extended unified theory of acceptance and use of technology. *Innovative Higher Education*, 49, 223–245. <https://doi.org/10.1007/s10755-023-09686-1>
- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10, 15. <https://doi.org/10.1186/s40561-023-00237-x>

- Yu, C., Yan, J., & Cai, N. (2024). ChatGPT in higher education: Factors influencing ChatGPT user satisfaction and continued use intention. *Frontiers in Education*, 9, 1354929. <https://doi.org/10.3389/feduc.2024.1354929>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *International Journal of Educational Technology in Higher Education*, 16, 39. <https://doi.org/10.1186/s41239-019-0171-0>