Generative AI in University Mathematics: Attitudes and Academic-Leisure Use Patterns

Hassan Hossein-Mohand[1], Hossein Hossein-Mohand[1], Manuel García-Alonso[2] and María del Carmen Olmos-Gómez[1]
[1]University of Granada, Spain
[2]UNED (National University of Distance Education), Spain

hassan.h.m@ugr.es
hossein.h.m@ugr.es (corresponding author)
mangarcia@melilla.uned.es
mcolmos@ugr.es

https://doi.org/10.34190/ejel.24.4.4850

Abstract: The rapid spread of generative artificial intelligence has reshaped university e-learning, yet its incorporation into mathematics learning remains constrained by disciplinary demands for precision, justification, and epistemic scrutiny. This study examined the relationship between socio-demographic variables, access conditions, and attitudes towards AI in mathematics, and analyzed how these factors were associated with reported use of AI tools in two differentiated contexts: academic learning and leisure. A quantitative, observational, non-experimental, cross-sectional study was conducted with 869 students from the University of Granada across the Melilla, Ceuta, and Granada campuses. Data were collected through an online questionnaire that included the IAMAT scale and two dichotomous indicators of AI use. Inferential analyses were estimated with between 834 and 846 complete cases, depending on the procedure. The IAMAT scale showed adequate internal consistency (α = .796), high sampling adequacy (KMO = .888), and an interpretable two-factor structure that distinguished between usefulness/confidence and uncertainty/errors. AI use was more frequent in mathematics learning (62.6%) than in leisure (28.3%). Logistic regression models indicated that positive attitudes towards AI in mathematics increased the likelihood of use in both contexts. In academic use, older age, lack of Wi-Fi access, and membership of the Melilla campus were associated with a lower probability of use, whereas participation in voluntary work was associated with a higher probability. In leisure use, women showed a lower probability of use than men. In addition, K-means clustering identified six differentiated profiles defined by age, perceived usefulness, and distrust. These profiles discriminated academic use significantly, but not leisure use. The findings suggest that the adoption of AI in mathematics cannot be reduced to technological access alone, because it is also shaped by domain-specific cognitive and affective dispositions. From an e-learning perspective, these findings contribute to AI-supported mathematics education by showing that digital learning environments should move beyond mere access to GenAI and embed critical AI literacy, mathematical verification criteria, and reasoning-oriented tasks in which students explain, check, and revise AI-generated solutions rather than delegate the full intellectual workload to the tool.

Keywords: Generative AI, Higher education, Mathematics learning, AI literacy, Attitudes towards AI, e-Learning

1. Introduction

The rapid diffusion of generative artificial intelligence across higher education has shifted the debate from technical availability to questions of epistemic responsibility, assessment design, and institutional governance (Bond et al., 2024; Jin et al., 2025). In e-learning environments, this shift is visible because large language models promise immediacy, personalisation, and conversational support, whilst also altering how students search for information, construct explanations, and delegate intellectual work to automated systems (Chan and Hu, 2023; Park, 2025; Rapanta et al., 2025). Mathematics learning sharpens these tensions because the value of an answer depends on conceptual precision, procedural validity, and the quality of the justification that supports it (Pepin, Buchholtz and Salinas-Hernández, 2025).

Within this domain, a plausible answer is not necessarily a mathematically valid one. Small symbolic deviations, omitted steps, or formally appealing but invalid inferences can change the status of a solution altogether (Yoon et al., 2024). For that reason, the educational significance of generative AI in mathematics cannot be inferred from generic technology-adoption models alone. What matters is how students judge the tool's usefulness, how they calibrate trust in its outputs, and under which conditions AI is incorporated into tasks such as searching for theory, translating verbal problems into mathematical language, checking solutions, or generating practice (Bewersdorff et al., 2025; Pham, Klaus and Bateh, 2025).

Cite this article: Hossein-Mohand, H. et al. 2026. "Generative AI in University Mathematics: Attitudes and Academic-Leisure Use Patterns", *Electronic Journal of e-Learning*, 24(4), pp 18-31, https://doi.org/10.34190/ejel.24.4.4850

Against this background, the present study addresses two linked empirical gaps. First, large-scale quantitative evidence remains limited on how multidimensional attitudes towards AI in mathematics are associated with reported use among university students, particularly in contexts connected with teacher education. Second, it remains unclear whether AI adoption in mathematics reflects a general orientation towards GenAI across everyday digital life or a more situated academic appropriation. For this reason, the study examines AI use in two contexts, academic learning and leisure. Leisure use is not treated as a direct mathematics-learning outcome, but as a theoretically useful contrast that helps determine whether AI use in mathematics is part of broader recreational engagement with GenAI or is specifically attached to study-related tasks and e-learning conditions. The study also identifies attitudinal profiles in order to determine whether different constellations of usefulness, distrust, and age are linked to differentiated patterns of AI use among university students. The theoretical framework is organised as a progression from the institutional level to the cognitive and pedagogical implications of generative AI in mathematics. It first addresses the move from institutional adoption to critical GenAI literacy, then examines the specific demands of mathematics education, before turning to attitudes, calibrated trust, and the relationship between pedagogical usefulness and cognitive risk. This progression provides the rationale for the study and clarifies why a domain-specific instrument such as IAMAT is analytically preferable to a generic measure of technology acceptance.

1.1 From Institutional Adoption to Critical GenAI Literacy

The arrival of generative AI in higher education can no longer be interpreted as a merely instrumental innovation (Bond et al., 2024; Jin et al., 2025). Comparative analyses of university policies suggest that institutional responses are increasingly structured around three interdependent concerns: academic integrity, redesign of teaching and assessment, and fairness of access. The central question is therefore not simply whether students use GenAI, but under which normative conditions, with what degree of human supervision, and according to which epistemic expectations. Jin et al. (2025), in their analysis of adoption strategies across universities from six world regions, show that institutions are moving from defensive responses to more complex frameworks based on ethical-use guidance, authentic assessment, and targeted training for staff and students. At the same time, they note that substantial gaps remain in privacy protection, transparency, and equitable access.

Within this context, generative AI literacy has become a key concept. Park (2025) frames it as a cluster of interrelated competences that includes understanding how the tool works, using and applying it, evaluating and incorporating its outputs, addressing ethical implications, and adopting an informed attitudinal stance towards it. Rapanta et al. (2025) extend this argument by insisting that critical GenAI literacy must include epistemic, ethical, and relational dimensions. Students therefore need more than prompt-writing skills. They need criteria for judging the credibility of an answer, recognising bias and omission, and deciding when the tool supports learning and when it encourages cognitive delegation. This shift is consequential in teacher education, where the use of AI shapes not only immediate performance but also future professional conceptions of knowledge, validation, and teaching (Bernardi et al., 2025; Busuttil and Calleja, 2025).

1.2 The Specificity of Mathematics Education in Relation to Generative AI

Mathematics education introduces an additional layer of difficulty because the educational value of AI depends on mathematical fidelity rather than on linguistic fluency alone (Yoon et al., 2024). In other domains, a partially approximate answer may still be serviceable; in mathematics, incomplete justification, symbolic imprecision, or plausible but invalid reasoning may undermine the solution itself (Noster, Gerber and Siller, 2024). Pepin, Buchholtz and Salinas-Hernández (2025), in their scoping review of ChatGPT in mathematics education, show that these tools are already being used to support conceptual understanding, lesson design, assessment, personalised assistance, and collaboration. Yet the same review makes clear that such affordances coexist with errors, hallucinations, dependency, and ethical concerns, which means that human oversight is not optional but structural.

From a mathematics education perspective, generative AI should therefore be understood as a resource that reconfigures students' mathematical activity rather than as a neutral shortcut to information. Kock, Salinas-Hernández and Pepin (2025) show that engineering students develop initial schemes of use for ChatGPT when they try to understand concepts, explore alternative modelling strategies, translate mathematical models into code, and optimise it. Contel and Cusi (2025) likewise report that GPT-4 can scaffold metacognitive processes during problem solving, especially planning, monitoring, and revision, provided that interaction with the system is organised in reflective ways. These contributions converge on an important point for the present study: the educational potential of GenAI is activated not by exposure to the tool itself, but by patterns of use that preserve student agency, task regulation, and evaluation of the produced response.

Recent work in mathematics education also suggests that the crucial issue is not only what generative AI can do, but what forms of error it encourages students to normalise. Noster, Gerber and Siller (2024) found that pre-service teachers were exposed to both correct and incorrect ChatGPT outputs across all analysed tasks, and that the likelihood of submitting an incorrect answer increased when the model's initial answer was also incorrect. In the domain of mathematical proof, Yoon et al. (2024) showed that students' uses of GenAI are mediated by their conceptions of proof, their conceptions of AI, and their ethical considerations, leading the authors to argue that university work with proof must explicitly teach critical engagement with AI outputs rather than passive acceptance. In teacher education, Bernardi et al. (2025) and Busuttil and Calleja (2025) further suggest that future teachers may judge the pedagogical appeal of AI-generated proposals more easily than their mathematical correctness, which makes the distinction between usefulness, credibility, and fidelity especially important. Overall, this body of work identifies a common tension that frames the present study. GenAI can support mathematical activity when it scaffolds explanation, monitoring, practice, and revision, but it can also weaken learning when students treat plausible output as already verified knowledge. This tension justifies examining how students' mathematics-specific attitudes towards usefulness, confidence, and possible error are associated differently with academic and non-academic use.

1.3 Attitudes, Calibrated Trust, and Patterns of Use

Research on ChatGPT adoption in higher education shows that attitudes towards AI do not form a simple continuum running from enthusiasm to rejection. Pham, Klaus and Bateh (2025) report that students' intention to use ChatGPT depends chiefly on perceived usefulness and subjective norm, whilst ease of use and trust act more indirectly by reinforcing usefulness. Bewersdorff et al. (2025) complicate the picture by identifying differentiated student profiles, including AI advocates, cautious critics, and pragmatic observers, and by showing that use, interest, attitudes, and AI literacy jointly shape AI self-efficacy. These findings justify treating attitudes not as a decorative background variable, but as a potential organising principle of students' relations with AI.

In mathematics, however, attitudes need to be conceptualised more narrowly than generic acceptance, because AI-supported mathematical activity requires students to evaluate usefulness together with correctness, justification, and possible error (Pepin, Buchholtz and Salinas-Hernández, 2025). The relevant construct includes perceived usefulness, confidence, awareness of error, willingness to verify, and ethical judgement. This is why the notion of calibrated trust is especially productive. A student with stronger AI literacy is not one who always trusts the system, but one who can distinguish when an output may serve as a hint, when it requires external checking, and when it should be rejected because of inconsistency or insufficient mathematical justification. This distinction helps to interpret the dimensions captured by the IAMAT scale, whose items combine perceived benefits with reservations about reliability in mathematically specific tasks. Accordingly, comparing academic and leisure use makes it possible to examine whether these attitudes merely reflect a general orientation towards AI or whether they are more strongly tied to mathematically framed learning tasks.

1.4 Pedagogical Usefulness, Cognitive Risk, and the Rationale for the Study

A theoretically adequate account of generative AI in education must also consider the relation between use and actual learning. Deng et al. (2025), in their meta-analysis of experimental studies, report overall positive effects on academic achievement, affective-motivational outcomes, and the propensity for higher-order thinking, together with a reduction in perceived mental effort. Yet Bauer et al. (2025) warn that better task performance under AI assistance does not necessarily translate into durable learning, and they argue that AI can operate through substitution, augmentation, redefinition, or even inversion of the intended learning process. In mathematics, this warning is particularly salient because an apparently helpful support may displace processes that are constitutive of learning itself, such as justifying a procedure, validating a result, checking symbolic consistency, or sustaining a proof.

Recent empirical evidence supports this cautious stance. Stadler, Bannert and Sailer (2024) found that large language models can reduce perceived cognitive load whilst at the same time producing shallower student reasoning in scientific inquiry tasks. Bastani et al. (2025) show that access to generative AI may improve immediate performance in mathematics and nonetheless harm subsequent learning when the tool is used as an answer machine rather than as support for reasoning. Among pre-service mathematics teachers, Zhang et al. (2025) further report that AI dependency is negatively associated with higher-order competences, whilst AI literacy partly mitigates that risk. Taken together, these studies suggest that the educational issue is not whether AI helps, but under which instructional conditions it helps without eroding the intellectual work it is meant to support.

These considerations return to the empirical gap introduced at the beginning of the article. In mathematics education, a substantial share of the recent evidence remains small-scale, exploratory, qualitative, or task-specific studies, often focused on proof, problem solving, or lesson planning (Contel and Cusi, 2025). Less is known, in large university samples and among future teachers, about how multidimensional attitudes towards AI in mathematics relate to reported use across academic and non-academic contexts. The comparison between academic learning and leisure is therefore not intended to make leisure use a direct mathematics-learning outcome. Rather, it provides a contrast that helps determine whether AI adoption is primarily a general digital habit or a situated e-learning practice linked to study tasks, mathematical verification, and access conditions. The present study addresses that gap by treating generative AI not as a homogeneous block, but as a situated practice tied to specific mathematical tasks and shaped by perceived usefulness, confidence, awareness of error, and access conditions. This perspective makes it possible to move beyond generic questions of adoption towards a more precise understanding of how students relate to AI when learning mathematics.

1.5 Study Objectives

This conceptual framework leads to one general objective and three specific objectives. Because leisure use is included as a comparison context, the objectives distinguish general engagement with GenAI from its academic appropriation in mathematics e-learning. G.O: To analyse how socio-demographic variables, access conditions, and attitudes towards AI in mathematics are related to reported use of AI tools in two contexts, academic learning and leisure.

- S.O.1.- To describe the prevalence of AI-tool use in learning and leisure and to examine the basic psychometric evidence for the IAMAT scale in the analysed sample.
- S.O.2.- To estimate the relative weight of attitudes towards AI in mathematics, together with socio-demographic and access variables, on the probability of AI use in both contexts.
- S.O.3.- To identify differentiated attitudinal profiles and to analyse their association with academic and leisure use of AI.

2. Method

2.1 Study Design

A quantitative, observational, non-experimental, cross-sectional study was conducted using a self-administered questionnaire. The design was intended to examine associations between socio-demographic variables, access conditions, and attitudes towards generative AI in mathematics, and to estimate their relationship with reported AI use in two distinct contexts, academic learning and leisure. Because all variables were observed at a single point in time and no experimental manipulation or random assignment was involved, the study had a descriptive and explanatory-predictive scope. The organisation of the method section followed the reporting logic recommended by STROBE for cross-sectional observational studies (von Elm et al., 2007).

2.2 Participants and Context

The target population comprised students from the University of Granada across the Melilla, Ceuta, and Granada campuses. Participants were recruited through non-probability convenience sampling based on questionnaire dissemination in the ordinary university context and on students' voluntary participation. A total of 869 questionnaires were validly recorded. For inferential analyses, complete-case criteria were applied according to the variables included in each procedure, which resulted in effective sample sizes ranging from 834 to 846 participants. In the sample used for the main analyses (n = 846), 67.6% identified as women (n = 572) and 32.4% as men (n = 274). Mean age was 21.76 years (SD = 4.56). Campus distribution was as follows: Melilla (n = 592, 70.0%), Ceuta (n = 159, 18.8%), and Granada (n = 95, 11.2%). With regard to access conditions, 95.5% reported having Wi-Fi connection (n = 808). Perceived socio-economic level was predominantly middle (n = 720, 85.1%), followed by low (n = 104, 12.3%) and high (n = 22, 2.6%). In addition, 22.2% reported being in work (n = 188) and 12.4% reported involvement in voluntary work (n = 105). This composition provided a heterogeneous basis for analysing AI use by age, campus, and contextual conditions, although no claim of probabilistic representativeness is made.

2.3 Instruments and Variables

Data were collected through an online questionnaire administered via Google Forms. The instrument contained 26 items distributed across three blocks. The first gathered socio-demographic and contextual variables, namely gender, age, campus membership, availability of Wi-Fi, perceived socio-economic level, employment status, and participation in voluntary work. The second included two dichotomous criterion variables, use of AI in learning activities (IA_APR: 0 = no, 1 = yes) and use of AI in leisure or free-time activities (IA_OCIO: 0 = no, 1 = yes). These variables were conceptualised as indicators of reported prevalence of use. Their function was to identify whether AI tools had entered each context of activity, academic learning and leisure, while the IAMAT scale captured students' mathematics-specific attitudes towards possible uses of AI. The third block included the IAMAT scale, which comprised nine five-point Likert items (1 = strongly disagree; 5 = strongly agree) designed to measure attitudes towards the use of AI in mathematics.

The IAMAT scale was developed from the mathematics-related component of a broader questionnaire on university students' use and perceptions of artificial intelligence. In its construction, the items were intended to measure attitudes towards AI-mediated mathematical activity. Item content was therefore organised around mathematically situated tasks and judgements, including searching for theoretical content, solving problems, working with geometry, translating real situations into mathematical language, checking solutions, generating practice exercises, and evaluating reliability.

Before administration, the questionnaire underwent a content-validation process based on expert judgement using a three-round Delphi procedure (Hsu and Sandford, 2007; Lynn, 1986). The expert panel consisted of 15 university lecturers holding a PhD and representing complementary areas related to the construct under study: mathematics and physics (five experts), pedagogy and research methods (six experts), and social sciences and linguistics (four experts). In each round, the experts assessed the appropriateness, relevance, and coherence of the items using a three-point scale ranging from 1 (not appropriate) to 3 (fully appropriate). The items were progressively refined in response to the experts' comments. These refinements involved wording and clarity adjustments, but no item was removed, because all items reached the predefined consensus threshold by the final round. The process ended with a 98% agreement rate, above the 90% criterion established for retaining the items. No separate pilot study with an independent subsample was conducted. Accordingly, the Delphi process functioned as the pre-fieldwork refinement stage, whereas the internal consistency analysis and exploratory factor analysis reported below provide initial empirical evidence of the scale's psychometric behaviour in the final analytical sample.

IAMAT assessed three complementary conceptual cores: perceived usefulness for specific mathematical tasks, confidence in AI-generated responses, and awareness of uncertainty or possible error. Substantively, the items covered searching for theoretical content, solving problems, support in geometry, translating real situations into mathematical language, checking solutions, generating exercises for practice, and judging reliability. Items 7 and 8 were phrased in the negative direction and were treated analytically as indicators of caution about error and reliability. Accordingly, higher values on IAMAT_DESCONF indicate greater distrust or stronger awareness of possible error. It should be noted that calibrated trust was used as an interpretive concept in the discussion to make sense of the coexistence of perceived usefulness, confidence, and awareness of possible error. The full wording of the items and their analytic orientation are provided in the Supplementary Material (Table S1).

For descriptive purposes and profile characterisation, three continuous indicators were derived from the scale: IAMAT_MEAN, computed as the mean of the nine items as a broad summary index; IAMAT_UTIL, computed as the mean of Items 1 to 6 and 9 so as to capture usefulness and confidence; and IAMAT_DESCONF, computed as the mean of Items 7 and 8 so as to represent distrust and concern about error. These composite indices were useful for description and clustering. For the logistic regression models, however, the composite indices were not entered simultaneously because of their conceptual overlap. Instead, standardised factor scores derived from the exploratory factor analysis were used as continuous predictors in order to represent attitudinal dimensions more parsimoniously and reduce redundancy between predictors.

2.4 Procedure

Data collection was conducted online. The questionnaire link was disseminated to the target population within ordinary university activity. The opening screen informed participants about the aim of the study, the voluntary nature of participation, the right to discontinue at any time, and the guarantees of confidentiality and anonymity. No direct personal identifiers were requested, and access to the questionnaire was conditional upon acceptance of informed consent. Once data collection had finished, the database was cleaned prior to analysis through the review of codings, consistency checks, and reverse coding of the negatively worded IAMAT items.

2.5 Data Analysis

Analyses were carried out with IBM SPSS Statistics 25.0. In a first phase, descriptive statistics were calculated for all variables. Absolute and relative frequencies were obtained for categorical variables, and means and standard deviations for continuous variables. Multivariate analyses were estimated using complete cases for the variables involved in each model.

In a second phase, the psychometric behaviour of the IAMAT scale was examined. Internal consistency was estimated by means of Cronbach's alpha and inspection of corrected item-total correlations. This procedure was chosen because it provides a useful first indicator of internal homogeneity in short survey scales, although its interpretation was made cautiously and always in connection with the dimensional structure of the instrument (Tavakol and Dennick, 2011). An exploratory factor analysis was then conducted using principal axis factoring with Oblimin oblique rotation. Although IAMAT was constructed around theoretically defined conceptual cores, a confirmatory approach was not prioritized at this stage because the scale had not yet been independently validated in this population and no separate validation sample was available. The analysis was therefore conceived as an initial examination of internal structure rather than as a strict test of a previously confirmed measurement model. Principal axis factoring was selected because the focus was on common variance, and Oblimin rotation was used because the underlying attitudinal dimensions were theoretically expected to correlate. Matrix factorability was checked through the Kaiser-Meyer-Olkin index and Bartlett's test of sphericity. Factor retention was guided by eigenvalues above 1, the loading pattern, and the conceptual coherence of the solution, following good-practice recommendations for exploratory validation studies (Watkins, 2018). Regression-based factor scores were saved for subsequent use as continuous predictors. Item-level psychometric diagnostics are reported in the Supplementary Material (Table S2).

In a third phase, two independent binary logistic regression models were estimated, one for IA_APR and one for IA_OCIO. In both models, the dependent variable was coded as 0 = no use and 1 = use. Age and the attitudinal factor scores were entered in a first block, and socio-demographic and contextual variables were entered in a second block, namely gender, availability of Wi-Fi, level of studies, campus, perceived socio-economic level, employment status, and voluntary work. Interpretation was based on B coefficients, Wald statistics, odds ratios with 95% confidence intervals, and overall model fit indicators, including the likelihood-ratio chi-square, Nagelkerke's R^2, and the Hosmer-Lemeshow test. The number of observed events in both criterion variables was also sufficient relative to the number of predictors considered, which favours estimation stability in logistic models of this complexity (Peduzzi et al., 1996). As an additional sensitivity analysis, and in response to the possibility that usefulness/confidence and uncertainty/errors might operate jointly, the FAC1_1 × FAC2_1 interaction term was added to both logistic models. This analysis assessed whether the association between usefulness/confidence and AI use varied according to the level of uncertainty/errors. The main-effects models were retained as the primary specification unless the interaction term produced a statistically and substantively meaningful improvement in model fit or interpretation.

In a fourth phase, segmentation techniques were applied in order to identify attitudinal profiles. The clustering variables were selected on both empirical and conceptual grounds. FAC1_1 and FAC2_1 represented the two empirically derived dimensions of IAMAT, namely usefulness/confidence and uncertainty/errors. IAMAT_MEAN was retained as a broad indicator of overall attitudinal intensity, and age was included as a continuous biographical variable because the adoption of AI in academic learning was expected to vary across student trajectories and age groups. The outcome variables, IA_APR and IA_OCIO, were deliberately excluded from the clustering procedure in order to avoid circularity, because the purpose was to examine afterwards whether independently derived profiles were associated with AI use. Other socio-demographic and contextual variables were also not used to define the clusters, since the aim was to obtain attitudinal profiles rather than groups driven primarily by gender, campus, or access conditions.

Preliminary solutions were explored through TwoStep clustering and K-means solutions were then estimated. A six-cluster solution was finally retained because it offered the best balance between empirical separation, stability across exploratory procedures, and substantive interpretability. Solutions with fewer clusters merged theoretically distinct configurations, such as students with high perceived usefulness but different levels of distrust, whereas more fragmented solutions produced very small or substantively redundant groups. The retained solution was interpreted from an explicitly analytical and descriptive logic, with the aim of capturing internal heterogeneity in the sample rather than postulating closed or exhaustive typologies. As an additional quality criterion, the classical literature on the interpretation and validation of non-hierarchical partitions was considered, particularly in relation to compactness and separation between groups (Rousseeuw, 1987).

Finally, the associations between the resulting profiles and AI use in learning and leisure were examined through contingency tables and Pearson's chi-square tests, with Cramér's V reported as an effect-size index. Statistical significance was set at $p < .05$ throughout.

2.6 Ethical Considerations

Participation was voluntary and based on informed consent. Anonymity of responses and confidentiality of information were guaranteed throughout the study. The research received a favourable opinion from the University of Granada Research Ethics Committee (ref. 4518/CEIH/2024, 16 September 2024). The approval confirmed compliance with the applicable ethical principles and with current personal-data protection regulations. Data were used exclusively for scientific purposes and were analysed and reported in aggregated form.

3. Results

3.1 Prevalence of Use and Psychometric Evidence for the IAMAT Scale

In relation to S.O.1, inferential analyses were estimated using complete cases, with n = 846 for the psychometric assessment and the model of AI use in learning, and n = 834 for the model of AI use in leisure. Overall, 530 students (62.6%) reported using AI tools for learning purposes, whereas 236 (28.3%) reported AI use in leisure or free time. Academic use was therefore more than twice as frequent as recreational use, which locates the phenomenon primarily in practices oriented towards task completion and study support rather than in undifferentiated general consumption.

The evidence gathered for this first objective also showed that the IAMAT scale has a sufficient psychometric basis for the subsequent analyses. Sampling adequacy was high (KMO = .888) and Bartlett's test of sphericity was significant, $\chi^2(36) = 4528.13$, $p < .001$. Principal axis factoring with Oblimin rotation identified two correlated dimensions that explained 62.90% of the common variance. As shown in Table 1, the first factor concentrated the applications with the clearest instrumental value for mathematical work, with particularly high loadings for problem solving (.897), geometry (.858), translation of verbal situations into mathematical language (.803), and checking solutions (.798). Generation of practice problems (.661) and general confidence in AI responses (.679) also loaded on this factor, albeit with slightly less weight. The second factor brought together the two indicators of epistemic caution associated with error and uncertainty, with loadings of .818 and .746. The correlation between factors was negative and moderate (r = -.285), which indicates that usefulness and caution do not operate as strictly opposite poles. At the descriptive level, item means fell within a moderate range (2.43-2.94). The highest mean was observed for checking solutions, whereas the lowest corresponded to the use of AI for geometry tasks. Internal consistency of the full scale was adequate (α = .796). Full item-level diagnostics are provided in the Supplementary Material (Table S2).

Table 1: Rotated factor matrix, communalities, and descriptive statistics for IAMAT items

Abbreviated item	M	SD	h²	F1	F2
Searching for theory	2.86	1.38	.664	.801	
Problem solving	2.70	1.37	.779	.897	
Geometry	2.43	1.33	.722	.858	
Verbal problems to mathematical language	2.67	1.31	.675	.803	
Checking solutions	2.94	1.36	.663	.798	
Generating practice problems	2.65	1.47	.473	.661	
Potential errors in mathematics	2.74	1.37	.711		.818
Reservations about reliability	2.80	1.31	.539		.746
Confidence in responses	2.67	1.19	.434	.679	

Note. Extraction by principal axis factoring and Oblimin rotation. Blank cells indicate the absence of a salient loading in the reported factor. KMO = .888; Bartlett's $\chi^2(36) = 4528.13$, $p < .001$; explained variance = 62.90%; α = .796; correlation between factors = -.285. F1 = usefulness/confidence; F2 = uncertainty/errors.

3.2 Predictors of AI Use in Learning and Leisure

To address S.O.2, Table 2 reports the final logistic models for both contexts of use. The model for IA_APR was significant, $\chi^2(13) = 82.61$, $p < .001$, showed moderate explanatory power (Nagelkerke $R^2 = .127$), and fitted the data adequately (Hosmer-Lemeshow, $p = .529$). The IA_OCIO model was also significant, $\chi^2(13) = 32.65$, $p = .002$, although with substantially lower explained variance (Nagelkerke $R^2 = .055$; Hosmer-Lemeshow, $p = .507$). Taken together, these values suggest that the considered predictors captured academic use more precisely than leisure use.

The most consistent pattern was the effect of positive attitudes towards AI in mathematics, which increased the probability of use in both domains: OR = 1.37 for learning and OR = 1.38 for leisure. By contrast, the uncertainty/errors factor did not reach statistical significance in either model. This result is theoretically important because it suggests that adoption depends less on eliminating reservations about reliability than on perceiving the tool as functionally valuable for mathematically relevant purposes.

Additional sensitivity models including the FAC1_1 × FAC2_1 interaction did not provide evidence that the effect of usefulness/confidence on AI use varied significantly according to uncertainty/errors. For academic AI use, the interaction term was not significant, B = .065, OR = 1.067, 95% CI [.904, 1.260], $p = .444$, and the increase in model fit was not significant, $\chi^2(1) = .586$, $p = .444$. For leisure AI use, the interaction term was also not significant, B = .112, OR = 1.119, 95% CI [.943, 1.327], $p = .198$, with no significant improvement in model fit, $\chi^2(1) = 1.676$, $p = .195$. The main-effects models were therefore retained as the primary specification. This result suggests that the relatively high academic use observed among distrustful moderates should be interpreted as a descriptive profile pattern rather than as evidence of a statistically confirmed interaction between the two attitudinal dimensions.

From that shared pattern onwards, the models diverged. In learning, age showed an inverse effect (OR = 0.93), so that each additional year of age was associated with a lower probability of use. Lack of Wi-Fi reduced the probability of academic use by roughly one half (OR = 0.47), and the contrast Melilla versus Ceuta was also significant (OR = 0.44). Given the uneven distribution of the sample across campuses, this contrast should be read as a sample-specific association rather than as evidence of a general campus-level effect. In addition, participation in voluntary work was associated with a higher probability of academic use (OR = 1.62). In leisure, by contrast, the only socio-demographic variable with a robust effect was gender: women showed a lower probability of recreational use than men (OR = 0.59). The full model coefficients are reported in the Supplementary Material (Tables S3 and S4).

Table 2: Logistic regression models for AI use in learning and leisure

Predictor	IA_APR OR [95% CI]	p	IA_OCIO OR [95% CI]	p
F1. Positive attitudes	1.37 [1.16, 1.62]	< .001	1.38 [1.16, 1.63]	< .001
F2. Uncertainty/errors	0.93 [0.78, 1.11]	.438	0.95 [0.79, 1.15]	.601
Age	0.93 [0.90, 0.96]	< .001	1.00 [0.96, 1.03]	.875
Gender (women vs men)	1.11 [0.80, 1.54]	.532	0.59 [0.42, 0.82]	.002
Wi-Fi (no vs yes)	0.47 [0.23, 0.94]	.032	0.49 [0.20, 1.22]	.125
Level of studies (global test)		.195		.970
Campus (global test)		< .001		.315
Melilla vs Ceuta	0.44 [0.29, 0.68]	< .001		
Socio-economic level (global test)		.014		.731
Employment (yes vs no)	1.35 [0.94, 1.95]	.103	1.02 [0.69, 1.49]	.940
Voluntary work (yes vs no)	1.62 [1.04, 2.53]	.034	0.65 [0.41, 1.04]	.074

Note. Odds ratios (ORs) and 95% confidence intervals are reported. Rows labelled global test correspond to Wald tests for polytomous predictors. For dichotomous predictors, the relevant contrasts are indicated directly in the row labels. Model statistics: IA_APR, $\chi^2(13) = 82.61$, Nagelkerke's $R^2 = .127$, Hosmer-Lemeshow $p = .529$; IA_OCIO, $\chi^2(13) = 32.65$, Nagelkerke's $R^2 = .055$, Hosmer-Lemeshow $p = .507$.

3.3 Attitudinal Profiles and Associations with AI Use

In line with S.O.3, K-means clustering identified six differentiated profiles based on FAC1_1, FAC2_1, IAMAT_MEAN, and age. Table 3 shows that most of the sample was concentrated in three young or early-adult clusters: distrustful moderates (39.0%), young sceptics (24.5%), and young enthusiasts (22.7%). By contrast, distrustful adults (2.4%) and older ambivalent students (1.3%) were very small edge profiles. They are retained in the table for transparency and to describe the full clustering solution, but they should not be used as a basis for substantive generalisation. The main interpretation therefore focuses on the larger and more stable profiles, especially young enthusiasts, young sceptics, distrustful moderates, and moderate adults.

Comparison across the larger profiles revealed a pattern that is not obvious at first sight: greater caution did not necessarily imply lower academic use. The distrustful moderates profile, which comprised 39.0% of the sample, showed a relatively high score in IAMAT_DESCONF (2.91) and still reported academic use above the sample average. This descriptive pattern reinforces the finding from the logistic models, namely that perceived usefulness discriminates adoption more clearly than the mere reduction of reservations about reliability. Patterns observed in the two smallest profiles are reported descriptively only and should be treated as exploratory.

Associations between cluster membership and AI use were not equivalent across the two contexts. The relationship with academic use was significant, $\chi^2(5) = 40.19$, $p < .001$, with a moderate effect size (Cramér's V = .22). The highest percentage of IA_APR was observed in young enthusiasts (76.6%), followed by distrustful moderates (63.6%) and young sceptics (59.4%); the lowest value corresponded to older ambivalent students (18.2%). By contrast, the association with leisure use did not reach significance, $\chi^2(5) = 7.17$, $p = .209$, and the effect size was small (V = .09). Variation in IA_OCIO was also narrower overall, ranging from 20.0% to 36.4%.

Table 3: Characterisation of cluster profiles in attitudes towards AI and age

Cluster	n (%)	IAMAT_MEAN M (SD)	IAMAT_UTIL M (SD)	IAMAT_DESCONF M (SD)	Age M (SD)
Moderate adults	86 (10.2)	2.66 (0.80)	2.67 (1.12)	2.60 (1.19)	26.66 (1.59)
Distrustful moderates	330 (39.0)	2.57 (0.83)	2.48 (1.10)	2.91 (1.26)	22.32 (1.04)
Young enthusiasts	192 (22.7)	3.51 (0.51)	3.80 (0.63)	2.52 (1.07)	19.31 (0.99)
Young sceptics	207 (24.5)	2.28 (0.54)	2.12 (0.64)	2.84 (1.16)	18.46 (0.96)
Distrustful adults	20 (2.4)	2.33 (0.56)	2.13 (0.84)	3.03 (1.32)	34.95 (3.32)
Older ambivalent students	11 (1.3)	2.53 (1.06)	2.56 (1.53)	2.41 (1.55)	47.36 (3.64)

Note. Clusters were obtained through K-means with k = 6 on FAC1_1, FAC2_1, IAMAT_MEAN, and age. Valid N = 846.

Figure 1 synthesises the gap between academic and leisure use across clusters. In nearly all profiles, academic use clearly exceeded recreational use, with especially wide gaps among young enthusiasts (41.3 percentage points) and distrustful moderates (36.5 points). The only apparent exception was the profile of older ambivalent students, in which leisure use (36.4%) exceeded academic use (18.2%). However, because this profile included only 11 participants, this reversal is reported for transparency but is not interpreted substantively. The more robust conclusion is that, in the larger profiles, attitudinal segmentation discriminated academic incorporation of AI more effectively than recreational use. Observed counts and row percentages are reported in the Supplementary Material (Table S5).

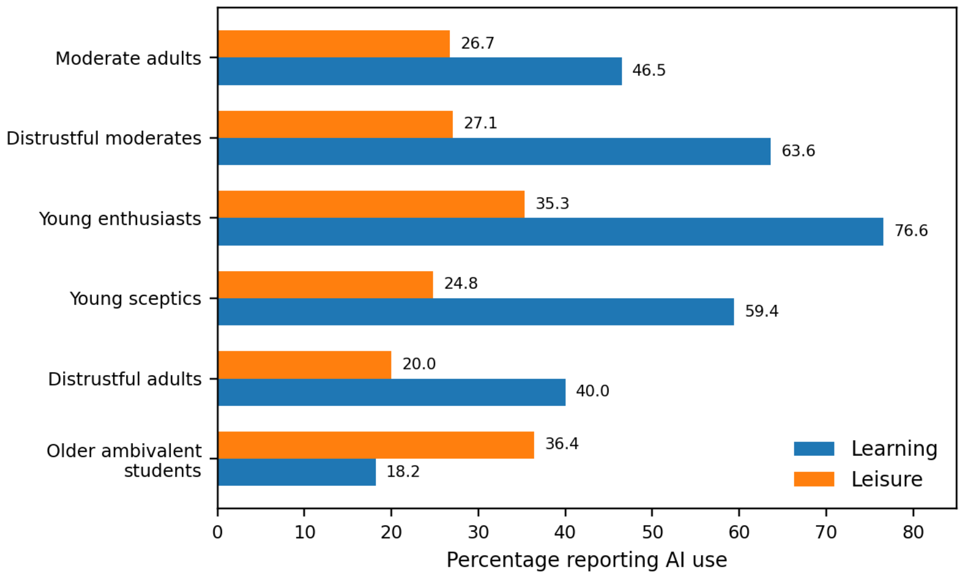

Figure 1: Reported AI use in learning and leisure by cluster

Note. Percentages represent students reporting AI use within each cluster. The association between cluster membership and IA_APR was significant, $\chi^2(5) = 40.19$, $p < .001$, $V = .22$, whereas the association between cluster membership and IA_OCIO did not reach significance, $\chi^2(5) = 7.17$, $p = .209$, $V = .09$.

4. Discussion

4.1 Meaning of the Main Findings

Taken together, the findings that respond to G.O. and, more specifically, to S.O.1 and S.O.2 show that the incorporation of generative AI into university mathematics learning does not follow a binary logic of acceptance versus rejection. What emerges instead is a pattern of selective adoption shaped by the value students attribute to the tool for specific academic purposes. The fact that use in learning substantially exceeds use in leisure suggests that, in this sample, AI is not functioning merely as a general-purpose technological novelty. It is being mobilised above all as an instrumental resource tied to concrete study needs, particularly searching for theory, checking solutions, and generating practice. This reading is consistent with recent work in higher education showing that student adoption of GenAI tends to combine pragmatic openness with persistent concerns about accuracy, privacy, integrity, and learning quality (Bond et al., 2024; Chan and Hu, 2023; Aldreabi et al., 2025).

Within that broader pattern, the most relevant result of the explanatory models is that positive attitudes towards AI in mathematics increased the probability of use in both learning and leisure, whereas the uncertainty/errors dimension was not statistically significant. This absence of effect should not be interpreted as evidence that critical judgement is irrelevant, but rather as a sign that recognising that AI can be wrong does not necessarily prevent students from using it. In a discipline such as mathematics, where checking is itself part of the learning activity, it is plausible that some students combine frequent use with active monitoring of error. The pattern is compatible with, but does not directly measure, a form of calibrated trust. In other words, some students may use AI while remaining aware of possible error and the need for verification. This interpretation should be treated cautiously because calibrated trust was inferred from the coexistence of perceived usefulness, reported use, and awareness of uncertainty, rather than measured as an independent construct. Even so, the pattern is consistent with recent arguments on critical GenAI literacy and with reviews specific to mathematics education that call for reflective, verifiable, and pedagogically purposeful uses of AI (Pepin et al., 2025; Rapanta et al., 2025).

4.2 Academic Versus Leisure Use: Interpretive Scope of the Contrast

The contrast between IA_APR and IA_OCIO should be understood as an analytical boundary test rather than as an attempt to treat leisure use as a direct mathematics-learning outcome. Its contribution lies in helping to distinguish general engagement with GenAI from its situated academic appropriation in mathematics e-learning. Although positive attitudes retained a stable effect in both models, academic use was much more sensitive to contextual and biographical variables than leisure use. The learning model retained age, availability of Wi-Fi, campus, and voluntary work, whereas the leisure model retained gender as its only robust socio-demographic predictor and showed much lower explanatory power overall. This asymmetry is informative: if the same predictors had explained both contexts with similar strength, the results would have pointed mainly to general digital engagement with AI. Instead, the weaker leisure model helps delimit the specifically academic pattern, suggesting that AI use in mathematics is more closely linked to study tasks, access conditions, and e-learning opportunity structures.

These contextual associations should be interpreted cautiously. The negative effect of age on academic use points to greater incorporation of AI into study routines among younger students, perhaps because of more recent exposure to these tools or because conversational interfaces feel closer to their ordinary digital practices. At the same time, the lower probability of use among students without Wi-Fi reminds us that the integration of AI continues to depend on basic material infrastructures and that the so-called GenAI divide cannot be detached from access conditions (Beckman et al., 2025). The campus contrast requires particular caution because the sample was unevenly distributed across sites, with a clear predominance of Melilla. It should therefore be interpreted as a contextual association within this sample, not as evidence that one campus is intrinsically more or less favourable to AI use. Differences in degree composition, teaching exposure, local digital cultures, or unmeasured institutional conditions may partly account for the observed pattern and should be examined in future stratified studies.

The positive association between voluntary work and academic use may reflect broader dispositions towards agency, social participation, or openness to innovation, although that interpretation requires further testing. In leisure, the lower probability of use among women is compatible with studies reporting demographic differences in perceived readiness, comfort, and benefit when students engage with generative AI (Maxwell et al., 2025). Yet the fact that this gender effect did not appear in the academic model is equally revealing, because it suggests that study-related uses may be less shaped by that gap than more general recreational engagement.

4.3 Implications for Mathematics Education and e-Learning

The findings related to S.O.3 reinforce the view that there is no single way of relating to AI in mathematics. The analytical value of the cluster profiles is diagnostic rather than taxonomic: they should not be read as fixed student types, but as configurations that help identify how perceived usefulness, distrust, and age combine in relation to academic AI use. In this respect, the distrustful moderates profile is especially informative because it shows that awareness of error can coexist with above-average academic use. This does not prove calibrated trust, but it suggests that distrust may sometimes operate as a verification-oriented stance rather than as a simple barrier to adoption. The small edge profiles are reported for completeness but are not used as a basis for substantive interpretation. Overall, the fact that profiles discriminated academic use more clearly than leisure use supports the domain relevance of IAMAT, attitudes towards AI in mathematics appear more informative when the activity is mathematically and academically framed.From the perspective of mathematics education and e-learning, the central implication is not that universities should increase AI use indiscriminately, but that they should improve its epistemic and pedagogical quality. A substantial share of students already use these tools for searching for theory, checking solutions, and practising, yet that does not by itself guarantee mathematically robust learning. Recent literature repeatedly warns that AI can support understanding, immediate feedback, and self-regulation whilst also fostering dependency, premature closure of reasoning, and weakened justification when it is used as a supplier of finished answers (Pepin et al., 2025; Bernardi et al., 2025).

For that reason, teacher education and mathematics e-learning should move the focus away from mere technical prompt proficiency and towards practices of verification and argumentation. Productive task designs might require students to compare human and AI-generated solutions, identify conceptual errors, request alternative explanations, reconstruct omitted steps, justify why an answer is acceptable, or discuss the limits of the tool according to the nature of the task. Such an orientation aligns closely with critical GenAI literacy (Park, 2025; Rapanta et al., 2025), because it treats AI not as an autonomous problem solver but as an object of disciplinary scrutiny.

In practical terms, the findings suggest three complementary lines of action for AI-supported mathematics e-learning. First, educators should guide students towards verification-oriented use. Instead of asking only whether AI was used, learning tasks can require students to document the AI-generated response, identify possible errors or omissions, justify which parts are mathematically valid, and produce a final revised solution. Secondly, AI can be integrated into task design as an object of critique rather than as a shortcut to the answer. For example, students may compare a human solution with an AI-generated solution, detect invalid inferences, reconstruct missing steps, or explain why two apparently plausible solutions are not equivalent. Thirdly, institutions should support responsible use through access, training, and policy. The relevance of Wi-Fi access in the academic model suggests that AI integration cannot be separated from digital infrastructure, while the role of attitudes indicates that training should address not only technical prompting but also mathematical verification, ethical use, and limits of automation.

In initial teacher education this is especially important: future teachers will not only learn with AI, but will also have to decide when, how, and for what purposes to integrate it into their own teaching.

This requirement becomes even more pressing in the light of evidence on the ambivalent effects of generative AI on learning. Bastani et al. (2025) show that access to GenAI may improve immediate mathematics performance and simultaneously weaken subsequent learning when the tool substitutes for reasoning. Stadler, Bannert and Sailer (2024) likewise found that large language models reduce mental effort but can compromise depth of inquiry. In pre-service mathematics teachers, Zhang et al. (2025) reported that AI dependency is negatively associated with higher-order competences, whereas AI literacy partly buffers that risk. Read against that evidence, the present findings point to both an opportunity and a warning. The opportunity lies in the fact that students already recognise concrete, potentially productive uses of AI. The warning is that more academic use does not automatically amount to better educational integration.

Finally, although the questionnaire underwent expert-based content validation through a Delphi procedure, no separate pilot administration with an independent subsample was conducted. The psychometric results should therefore be understood as initial evidence obtained in the final analytical sample, and future research should replicate the IAMAT structure through confirmatory factor analysis in independent samples.

4.4 Limitations and Research Agenda

The study has several limitations that should shape interpretation of the findings. First, it is a cross-sectional study based on self-report, so it cannot establish causal direction and cannot show how attitudes and uses evolve over time. Secondly, the sample comes from a single university and was recruited by convenience, which restricts external generalisation. In addition, campus distribution was uneven, with a clear predominance of Melilla; therefore, campus contrasts should be interpreted as sample-specific contextual associations rather than as general campus-level effects. Thirdly, the dependent variables register only presence or absence of use, not its frequency, intensity, quality, specific purpose, or the concrete AI tool employed. They should therefore be interpreted as prevalence indicators rather than as measures of how students use AI. From a mathematics-education standpoint, using AI occasionally to check a procedure is not equivalent to using it to delegate the full solution of a task, and this distinction matters substantively. Future studies should include graded indicators of frequency, identify the specific AI tools used, and distinguish task-level purposes such as searching for theory, checking solutions, generating practice exercises, translating verbal problems into mathematical language, or obtaining complete answers. A more specific methodological limitation should also be noted. IAMAT includes usefulness and confidence items framed in relation to concrete mathematical tasks, which gives the instrument high domain relevance but also places it close to the very phenomenon of academic use that it later helps to explain. Part of the association between positive attitudes and AI use may therefore reflect conceptual proximity between predictor and criterion rather than a relation between fully independent constructs. This does not invalidate the findings, but it does suggest that future studies should separate more clearly prior beliefs about AI, accumulated experience of use, actual frequency of use, and effects on mathematical performance. Further work would also benefit from longitudinal designs, measures of solution quality or performance, logged interaction data from specific tools, and comparisons across degree programmes and universities.

5. Conclusions

In relation to G.O, the study shows that generative AI use in university mathematics learning follows a selective and situated logic that is more complex than a reading centred solely on technical availability would suggest. Positive attitudes towards AI in mathematics were associated with a higher probability of use in both learning and leisure, but academic use was clearly more frequent and more sensitive to age, access, and institutional context. The leisure-use findings contribute to this argument by functioning as a contrastive boundary: they show that the academic appropriation of AI in mathematics cannot be reduced to a general recreational orientation towards GenAI. Instead, it appears more closely linked to study-related tasks, e-learning conditions, and domain-specific attitudes. Educationally, the results support a route that avoids both uncritical permissiveness and defensive restriction. AI-supported mathematics learning should be designed around verification, explanation, justification, and revision. In practice, this means asking students not only to obtain AI-generated answers, but to examine their validity, detect errors, reconstruct missing reasoning, and decide when the tool supports learning and when it replaces the intellectual work that mathematics education seeks to cultivate. In initial teacher education this orientation is especially urgent, because future teachers will need to judge not only whether AI can be used, but when its use strengthens mathematical reasoning and when it undermines it. The central contribution of the study is therefore both theoretical and practical. Theoretically, it shows that AI adoption in mathematics is patterned by domain-specific attitudes, access conditions, and differentiated student profiles rather than by access alone or by generic technology acceptance. Practically, it indicates that e-learning policies and mathematics-education practices should integrate AI through guided, accountable, and verification-oriented tasks, supported by institutional training and equitable access. Recognising this heterogeneity is a necessary step if AI is to strengthen, rather than erode, mathematical reasoning in digital learning environments.

6. Supplementary Material

A separate supplementary file accompanies the article to increase reporting transparency without overloading the main manuscript. It includes the full wording and analytic orientation of the IAMAT items (Table S1), item-level psychometric diagnostics (Table S2), the complete coefficients of the logistic models for IA_APR and IA_OCIO (Tables S3 and S4), and the observed counts and percentages of AI use in learning and leisure by cluster (Table S5).

7. Declarations

AI Statement: Generative AI tools were used during manuscript preparation to support linguistic reorganisation and formal refinement of some draft sections. The selection of evidence, interpretation of results, argumentative decisions, and final revision of the manuscript remained the sole responsibility of the authors.

Ethics Statement: The study received a favourable opinion from the University of Granada Research Ethics Committee (ref. 4518/CEIH/2024, 16 September 2024). All participants gave informed consent before taking part, participation was voluntary, and data were processed anonymously and in aggregated form.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any personal, academic, or financial relationships that could be construed as a potential conflict of interest.

Author Contributions: Conceptualization, Hassan H.M. and Hossein H.M.; software, Hossein H.M.; formal analysis, Hossein H.M. and María del Carmen O.G.; document analysis/investigation, Hassan H.M.; data curation, Hassan H.M. and Hossein H.M.; writing – original draft, Hossein H.M.; writing – review and editing, Hassan H.M. and Manuel G.A.; supervision, Manuel G.A.; project coordination and funding acquisition, María del Carmen O.G.

Funding: This work was supported by the research project MEL-14-UGR24, *"Iluminando oportunidades interseccionales: aprendizaje para la mejora educativa y laboral del uso de la inteligencia artificial (IA) en jóvenes"* [*Illuminating intersectional opportunities: learning for educational and professional improvement through the use of artificial intelligence (AI) among young people*], funded under the 2024 UGR/Autonomous City of Melilla Research Projects programme.

References

Aldreabi, H., Dahdoul, N.K.S., Alhur, M., Alzboun, N. and Alsalhi, N.R., 2025. Determinants of student adoption of generative AI in higher education. Electronic Journal of e-Learning, 23(1), pp. 15-33. https://doi.org/10.34190/ejel.23.1.3599

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö. and Mariman, R., 2025. Generative AI without guardrails can harm learning: Evidence from high school mathematics. Proceedings of the National Academy of Sciences, 122(26), e2422633122. https://doi.org/10.1073/pnas.2422633122

Bauer, E., Greiff, S., Graesser, A.C., Holstein, K., Rummel, N., Hmelo-Silver, C.E., Aleven, V. and McNamara, D.S., 2025. Looking beyond the hype: Understanding the effects of AI on learning. Educational Psychology Review, 37, 45. https://doi.org/10.1007/s10648-025-10020-8

Beckman, K., Apps, T., Howard, S.K., Rogerson, C., Rogerson, A. and Tondeur, J., 2025. The GenAI divide among university students: A call for action. The Internet and Higher Education, 67, 101036. https://doi.org/10.1016/j.iheduc.2025.101036

Bernardi, M.L., Capone, R., Faggiano, E. and Rocha, H., 2025. Generative AI in mathematics education: Pre-service teachers' knowledge and implications for their professional development. International Journal of Mathematical Education in Science and Technology, 56(8), pp. 1513-1530. https://doi.org/10.1080/0020739X.2025.2490104

Bewersdorff, A., Hornberger, M., Nerdel, C. and Schiff, D.S., 2025. AI advocates and cautious critics: How AI attitudes, AI interest, use of AI, and AI literacy build university students' AI self-efficacy. Computers and Education: Artificial Intelligence, 8, 100340. https://doi.org/10.1016/j.caeai.2024.100340

Bond, M., Khosravi, H., de Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S.W. and Siemens, G., 2024. A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. International Journal of Educational Technology in Higher Education, 21, 4. https://doi.org/10.1186/s41239-023-00436-z

Busuttil, L. and Calleja, J., 2025. Teachers' beliefs and practices about the potential of ChatGPT in teaching mathematics in secondary schools. Digital Experiences in Mathematics Education, 11, pp. 140-166. https://doi.org/10.1007/s40751-024-00168-3

Chan, C.K.Y. and 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

Contel, F. and Cusi, A., 2025. Investigating the role of ChatGPT in supporting metacognitive processes during problem-solving activities. Digital Experiences in Mathematics Education, 11, pp. 167-191. https://doi.org/10.1007/s40751-024-00164-7

Deng, R., Jiang, M., Yu, X., Lu, Y. and Liu, S., 2025. Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. Computers & Education, 227, 105224. https://doi.org/10.1016/j.compedu.2024.105224

Hsu, C.-C. and Sandford, B.A., 2007. The Delphi technique: Making sense of consensus. Practical Assessment, Research and Evaluation, 12(10), pp. 1-8. https://doi.org/10.7275/pdz9-th90

Jin, Y., Yan, L., Echeverria, V., Gašević, D. and Martinez-Maldonado, R., 2025. Generative AI in higher education: A global perspective of institutional adoption policies and guidelines. Computers and Education: Artificial Intelligence, 8, 100348. https://doi.org/10.1016/j.caeai.2024.100348

Kock, Z.-J., Salinas-Hernández, U. and Pepin, B., 2025. Engineering students' initial use schemes of ChatGPT as an instrument for learning. Digital Experiences in Mathematics Education, 11, pp. 192-218. https://doi.org/10.1007/s40751-025-00169-w

Lynn, M.R., 1986. Determination and quantification of content validity. Nursing Research, 35(6), pp. 382-386 https://doi.org/10.1097/00006199-198611000-00017

Maxwell, D., Oyarzun, B., Kim, S. and Bong, J.Y., 2025. Generative AI in higher education: Demographic differences in student perceived readiness, benefits, and challenges. TechTrends, 69, pp. 1248-1259. https://doi.org/10.1007/s11528-025-01109-6

Noster, N., Gerber, S. and Siller, H.-S., 2024. Pre-service teachers' approaches in solving mathematics tasks with ChatGPT. Digital Experiences in Mathematics Education, 10, pp. 543-567. https://doi.org/10.1007/s40751-024-00155-8

Park, J., 2025. A systematic literature review of generative artificial intelligence literacy in schools. Computers and Education: Artificial Intelligence, 9, 100487. https://doi.org/10.1016/j.caeai.2025.100487

Peduzzi, P., Concato, J., Kemper, E., Holford, T.R. and Feinstein, A.R., 1996. A simulation study of the number of events per variable in logistic regression analysis. Journal of Clinical Epidemiology, 49(12), pp. 1373-1379. https://doi.org/10.1016/S0895-4356(96)00236-3

Pepin, B., Buchholtz, N. and Salinas-Hernández, U., 2025. A scoping survey of ChatGPT in mathematics education. Digital Experiences in Mathematics Education, 11, pp.9-41. https://doi.org/10.1007/s40751-025-00172-1

Pham, L., Klaus, T. and Bateh, J., 2025. Key factors influencing intention to use ChatGPT: An empirical study of U.S. students. Acta Psychologica, 260, 105516. https://doi.org/10.1016/j.actpsy.2025.105516

Rapanta, C., Bhatt, I., Bozkurt, A., Chubb, L.A., Erb, C., Forsler, I., Gravett, K., Koole, M., Lintner, T., Örtegren, A., Petricini, T., Rodgers, B., Webster, J., Xu, X., Christensen, I.-M.F., Dohn, N.B., Christensen, L.L.W., Zeivots, S. and Jandrić, P., 2025. Critical GenAI literacy: Postdigital configurations. Postdigital Science and Education, 7, pp. 1296-1333. https://doi.org/10.1007/s42438-025-00573-w

Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20, pp. 53-65. https://doi.org/10.1016/0377-0427(87)90125-7

Stadler, M., Bannert, M. and Sailer, M., 2024. Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. Computers in Human Behavior, 160, 108386. https://doi.org/10.1016/j.chb.2024.108386

Tavakol, M. and Dennick, R., 2011. Making sense of Cronbach's alpha. International Journal of Medical Education, 2, pp. 53-55. https://doi.org/10.5116/ijme.4dfb.8dfd

von Elm, E., Altman, D.G., Egger, M., Pocock, S.J., Gøtzsche, P.C. and Vandenbroucke, J.P., 2007. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. PLoS Medicine, 4(10), e296. https://doi.org/10.1371/journal.pmed.0040296

Watkins, M.W., 2018. Exploratory factor analysis: A guide to best practice. Journal of Black Psychology, 44(3), pp. 219-246. https://doi.org/10.1177/0095798418771807

Yoon, H., Hwang, J., Lee, K., Roh, K.H. and Kwon, O.N., 2024. Students' use of generative artificial intelligence for proving mathematical statements. ZDM - Mathematics Education, 56, pp. 1531-1551. https://doi.org/10.1007/s11858-024-01629-0

Zhang, D., Wijaya, T.T., Wang, Y., Su, M., Li, X. and Damayanti, N.W., 2025. Exploring the relationship between AI literacy, AI trust, AI dependency, and 21st century skills in preservice mathematics teachers. Scientific Reports, 15, 14281. https://doi.org/10.1038/s41598-025-99127-0