

Article

Generative AI Adoption and Perceived Academic Impact Among Undergraduate Students: A Cross-National Study in Indonesia, Tajikistan, and the United States

Vyacheslav Dushenkov ^{1,*}, Ernawati Sinaga ², Astri Rozanah Siregar ², Alamkhon Akhmedov ³ and Saidbeg Satorov ³

¹ Department of Natural Sciences, Hostos Community College CUNY, Bronx, NY 10451, USA

² Department of Biology, Universitas Nasional, Jalan Sawo Manila 61, Jakarta 12520, Indonesia; ernawatisinaga@unas.ac.id (E.S.); astri.siregar.rozanah@gmail.com (A.R.S.)

³ Medical—Social Institute of Tajikistan, Dushanbe 734042, Tajikistan; alamkhon.01021951@gmail.com (A.A.); satorov1955@gmail.com (S.S.)

* Correspondence: vdushenkov@hostos.cuny.edu

Abstract

Generative artificial intelligence (AI) is rapidly transforming higher education, yet comparative evidence from diverse international settings remains limited. This cross-national study examined undergraduate students' adoption and perceived academic impact of generative AI tools at institutions in Indonesia, Tajikistan, and the United States. A harmonized online survey was administered to 584 undergraduates during Spring 2026 (Indonesia $n = 235$; Tajikistan $n = 226$; United States $n = 123$). The survey assessed AI familiarity, adoption patterns, frequency and purpose of use, perceived academic impact, and factors encouraging AI engagement. AI adoption differed significantly across sites, with the highest adoption observed in Indonesia (84.3%), followed by Tajikistan (67.7%) and the United States (47.2%). These differences remained significant after adjustment for demographic and educational variables. Self-rated AI familiarity was the strongest predictor of both adoption and perceived academic benefit across all three settings. Differences across sites persisted even after accounting for participants' academic discipline. The perceived-impact scale demonstrated excellent internal consistency (Cronbach's $\alpha = 0.907$). Brainstorming, writing assistance, and broader task diversification were associated with higher perceived benefit. Exploratory analyses found limited evidence linking learning-style preferences to AI use patterns. The findings highlight the importance of institutional context, self-rated AI familiarity, and curriculum integration in shaping student engagement with generative AI.

Keywords: generative artificial intelligence; higher education; cross-national study; AI literacy; technology adoption; ChatGPT; perceived academic impact; educational technology



Academic Editors: Eleanor Dommett,
Neil Gordon and Xiyang Li

Received: 17 May 2026

Revised: 15 July 2026

Accepted: 17 July 2026

Published: 20 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Generative artificial intelligence (AI) has emerged as one of the most consequential technological developments in contemporary higher education. Since the public release of large language model (LLM) platforms such as ChatGPT, Gemini, Claude, and Perplexity AI, students have rapidly integrated generative AI into everyday academic practices, including studying, writing, brainstorming, coding, translation, and information retrieval. Adoption has outpaced the development of institutional policy, pedagogical guidance, and

empirical evidence, and undergraduate students are now among the earliest and most intensive adopters of these technologies worldwide.

The educational implications of this rapid adoption remain contested (Francis et al., 2025). Advocates contend that generative AI can enhance learning efficiency, creativity, accessibility, and personalization, and emerging evidence suggests it can improve brainstorming, drafting, and self-directed learning when integrated appropriately. Critics, however, raise concerns about academic integrity (Annas & Nugraha, 2026), overreliance on automated systems, bias amplification, misinformation, and the erosion of critical thinking (Tahir & Zahra, 2026; Jin et al., 2025). Recent scholarship has therefore moved beyond polarized debates toward more nuanced examinations of how students actually use generative AI in authentic settings (Perkins, 2023) and how institutional context shapes those behaviors.

Despite this rapidly expanding literature, important gaps remain. Most published research consists of single-institution studies from high-income English-speaking countries, often with a focus on computer science students (Zou et al., 2026), and cross-national comparative investigations remain rare (Jin et al., 2025; Marmolejo-Ramos et al., 2026). As a result, relatively little is known about how AI adoption differs across institutional contexts, linguistic environments, and educational cultures. This gap is especially pronounced for low- and middle-income settings and for regions underrepresented in educational technology research, including Central Asia and parts of Southeast Asia.

A growing body of research suggests that institutional and contextual factors may matter more than individual demographics in shaping educational AI adoption (Sergeeva et al., 2025). Faculty endorsement, policy clarity, technological accessibility, curriculum integration, and AI literacy training all appear to shape whether students view generative AI as academically legitimate and beneficial. Emerging frameworks, therefore, conceptualize AI adoption as a sociotechnical, ecosystem-level phenomenon arising from interactions among institutional governance, disciplinary culture, technological infrastructure, and learner characteristics rather than as an individual technology acceptance problem alone (Chiu, 2025; Dong et al., 2025).

A separate, more exploratory question concerns whether learners with different self-reported modality preferences preferentially adopt different categories of multimodal AI tools. Although the broader educational utility of learning-style frameworks such as VARK remains contested (Brennan & Barbon, 2026; Marwaha & Sharma, 2025), the proliferation of multimodal AI systems makes the alignment between modality preferences and AI tool selection an empirical question worth describing, even if not a primary focus of the present study (Pashler et al., 2008).

Conceptual Framework

The present study is guided by a conceptual framework that integrates three complementary perspectives on educational AI adoption. First, drawing on the Unified Theory of Acceptance and Use of Technology (UTAUT2) (Sergeeva et al., 2025), we treat adoption as a function of perceived usefulness, ease of use, and facilitating conditions—faculty endorsement, curriculum integration, and institutional policy—rather than individual attitudes alone. Second, we adopt an AI literacy lens, conceptualizing AI literacy as a multidimensional competency encompassing technical understanding, ethical awareness, critical evaluation, responsible use, and confidence in interacting with AI systems (Chiu, 2025; Dong et al., 2025). In the present study, this construct is partially operationalized through self-reported AI familiarity. Given that self-reported familiarity primarily taps into confidence and prior exposure rather than the full multidimensional construct, we regard it throughout as a partial indicator, or proxy, for AI literacy, rather than as an exhaustive mea-

sure of the construct as articulated in recent frameworks (Ng et al., 2021). Third, we situate adoption within a sociotechnical ecosystem in which institutional governance, disciplinary culture, technological infrastructure, and learner characteristics jointly shape whether and how students engage with generative AI. Together, these perspectives motivate the study's comparative, multi-institutional design and direct analytic attention to both individual-level predictors (familiarity, prior exposure) and contextual moderators (institution, discipline, language environment). Within this framework, learning-style preferences are treated as a descriptive, exploratory dimension rather than a core explanatory construct.

The present study addresses these gaps through a cross-national comparative investigation of undergraduate AI tool use across three higher education environments: Hostos Community College of the City University of New York (United States), Universitas Nasional (Indonesia), and the Medical–Social Institute of Tajikistan. These three settings were selected to maximize variation in the institutional and national conditions thought to shape generative AI adoption, rather than as a representative sample of their respective countries. They differ in institutional type (a U.S. open-access community college, an Indonesian private comprehensive university, and a Tajik specialized health-sciences institute), in the dominant language of instruction and survey administration (English; English within a primarily Indonesian-speaking environment; and Russian), and in national economic context (one high-income and two middle-income countries). They also span different higher-education systems, levels of digital infrastructure, and degrees of institutional AI policy maturity. This purposive, maximum-variation design allows the study to examine whether patterns of adoption and perceived impact recur across heterogeneous contexts or instead track specific institutional and linguistic conditions; consistent with this aim, the three sites proved to differ substantially in their realized AI tool ecosystems and in the factors students identified as encouraging further use (Section 3). The study pursues a hierarchy of analytic aims. The primary, confirmatory objectives are: (1) to characterize differences across the three institutional sites in AI adoption, frequency of use, and perceived educational impact; and (2) to identify the individual- and institutional-level factors most strongly associated with AI adoption and perceived benefit. A secondary objective is to evaluate the psychometric performance of a seven-item perceived-impact scale for use in future cross-institutional AI research. An exploratory objective is to describe whether students with different learning-style preferences (VARK) tend to gravitate toward different categories of AI tools; given the ongoing debate over the validity of learning-style frameworks, this analysis is treated as hypothesis-generating rather than confirmatory. This explicit hierarchy guides the interpretation of statistical results throughout the manuscript.

These aims correspond to four explicit research questions. RQ1 (confirmatory): How do AI adoption, frequency of use, and perceived academic impact differ across the three institutional sites? RQ2 (confirmatory): Which individual- and institutional-level factors are most strongly associated with AI adoption and perceived benefit? RQ3 (secondary): Does the seven-item perceived-impact scale show adequate psychometric performance for cross-institutional use? RQ4 (exploratory): Do students with different learning-style (VARK) preferences gravitate toward different categories of AI tools? RQ1 and RQ2 are confirmatory, RQ3 is secondary, and RQ4 is treated as exploratory and hypothesis-generating.

By integrating cross-national comparison, AI literacy analysis, and task-specific educational use within a single harmonized framework, the present investigation contributes new evidence to the emerging literature on generative AI in higher education. The inclusion of undergraduate populations from Indonesia and Tajikistan addresses a substantial geographical gap in existing research and offers insight into how generative AI adoption unfolds beyond the high-income Western contexts that currently dominate the literature (Marmolejo-Ramos et al., 2026).

2. Materials and Methods

2.1. Research Sites

This study employed a non-experimental, cross-sectional, anonymous online survey design to examine undergraduate students' adoption and perceived academic impact of generative artificial intelligence (AI) tools, and to explore associations between learning-style preferences and AI use. Data were collected in parallel at three partner institutions:

Hostos Community College, City University of New York (CUNY), Bronx, NY 10451, USA;

Universitas Nasional, Jalan Sawo Manila 61, Jakarta 12520, Indonesia;

Medical–Social Institute of Tajikistan, Dushanbe 734042, Republic of Tajikistan.

Each site used the same instrument and operational procedures, allowing pooled analysis and site-level comparisons. Because each national context is represented by a single institution, between-site differences cannot be attributed to the country *per se*; throughout this article, we therefore interpret them as differences across three institutional sites located in different national contexts, recognizing that such differences may reflect a combination of country, institution, disciplinary composition, language of instruction, local AI policy, and tool availability. Country labels are retained as shorthand for these institutional sites rather than as estimates of national-level effects. The study was conducted in accordance with the Declaration of Helsinki ([World Medical Association, 2025](#)).

2.2. Participants and Eligibility

Participants were undergraduate students enrolled at one of the three partner institutions at the time of survey completion. Inclusion criteria were (i) current enrolment as an undergraduate, and (ii) age ≥ 18 years. Eligibility was verified through a screening item at the start of the instrument; respondents who did not meet both criteria were exited automatically and no response was recorded. A non-probability convenience sampling strategy was used at all three sites. The final analytic dataset comprised 584 valid responses (Indonesia $n = 235$; Tajikistan $n = 226$; United States $n = 123$). Participation was voluntary and uncompensated. No academic incentive (e.g., extra credit) was tied to participation, and recruitment materials explicitly stated that the decision to participate or to withdraw would have no effect on grades, academic standing, or any relationship with course instructors. Respondents could skip any item or exit the survey at any time prior to submission; partially completed responses were not used to identify individuals.

2.3. Survey Instrument

Data were collected using a self-administered online questionnaire hosted on Microsoft Forms via the Hostos Microsoft 365 for Education tenant. The same instrument was deployed at all three sites. The survey was administered in English at the United States and Indonesian sites and in Russian at the Tajik site, where students are fluent in Russian. For the Tajik administration, the instrument was translated from English into Russian and back-translated into English by two bilingual members of the research team (V.D. and S.S.), both fluent in both languages, to verify semantic equivalence. Because no formal cognitive pretesting or local pilot was undertaken, this procedure affords semantic-equivalence assurance rather than full cross-cultural validation. At the Indonesian site, the instrument was administered in English to a predominantly non-native-English-speaking population. The implications of these features for cross-site comparability are considered in Section 4. The instrument consisted of three sections; the items used in the present analyses are described below.

2.3.1. Demographics and Background

Five items captured age group (eight ordered categories: <18, 18–21, 22–25, 26–30, 31–35, 36–40, 41–50, >50; respondents indicating <18 were excluded by the eligibility screen), first language (English, Spanish, Other), gender (Female, Male, Prefer not to say), the courses in which the respondent was currently enrolled (open-ended; used to classify discipline post hoc), and prior experience or formal education in computer science, data science, or AI-related fields (Yes/No).

2.3.2. AI Familiarity, Adoption, and Use

Six items captured (i) self-rated familiarity with the concept of AI on a 5-point scale (“Not at all familiar” to “Extremely familiar”); (ii) AI tool adoption status (Yes/No/Maybe); (iii) awareness or use of specific tools, with a multiple-response checklist (ChatGPT, Gemini, Copilot, Grammarly, QuillBot, Perplexity AI, DALL·E/Stable Diffusion, Other); (iv) free-text identification of the most frequently used tool; (v) frequency of use for academic or professional purposes (Daily, Weekly, Monthly, Rarely, Never); and (vi) types of tasks for which AI was used, as a multiple-response checklist of nine task categories (studying and learning; brainstorming; writing assistance; coding and programming support; data analysis or visualization; language learning; presentation and design support; administrative tasks; Other). A subsequent open-ended item invited respondents to elaborate on how AI had supported the task category they selected.

2.3.3. Perceived Impact and Usefulness

The perceived impact of AI on academic performance was assessed using 7 items, each scored on a 5-point ordinal scale (1 = Not at all, 2 = Slight impact, 3 = Moderate impact, 4 = Strong impact, 5 = Significant impact). The seven domains were: ability to learn and understand new material; efficiency in completing assignments; quality of writing and communication; creativity in problem-solving or projects; preparation for exams or assessments; ability to analyze and interpret data; and collaboration or group-work experience. A separate single-item global rating captured overall perceived usefulness of AI tools in academic work (5-point scale: “Extremely useful” to “Not at all useful”). A multiple-response item asked what would encourage greater AI use (six pre-specified categories plus “Other”).

2.3.4. Learning-Style Preferences

Learning-style preferences were assessed using the VARK Questionnaire version 9.2 (Visual, Auditory, Reading/Writing, Kinesthetic), administered in its standard 16-item multiple-response format. Respondents were instructed to select the option(s) that best matched their preferences for each scenario and to leave any item that did not apply blank. The four VARK scale scores were derived per the questionnaire’s published scoring rules. Because VARK profiles are non-exclusive (a respondent may endorse one or more modalities), each respondent’s profile was re-expressed for analysis as four binary modality-endorsement indicators (V, A, R, K), together with a count of modalities endorsed and a unimodal/multimodal status indicator. These two coding schemes—binary endorsement of each modality, and unimodal vs. multimodal profile—are used in the analyses reported below.

2.4. Data Collection Procedures

The survey was hosted on a password-protected Microsoft 365 for Education tenant accessible only to the principal investigator at the Hostos site, with parallel arrangements at the partner sites. Respondents accessed the survey via a hyperlink or QR code and

completed it in a single session of approximately 15–20 min at a time and place of their own choosing. The form was configured to collect responses anonymously: no names, e-mail addresses, IP addresses, or institutional identifiers were captured. Data collection was conducted during the Spring 2026 academic semester, with the survey open for approximately three to four weeks; the partner sites collected data using the same instrument across the same overall study period. No follow-up contact, repeated measures, or linked records were involved.

2.5. Data Management, Privacy, and Confidentiality

Upon survey closure, the raw response file was exported to Microsoft Excel and transferred to an encrypted, password-protected folder on a college-owned, firewall-protected desktop computer located in the principal investigator's locked office. No research data were stored on personal devices, removable media, or non-institutional cloud services. Access to the dataset was limited to the principal investigator and approved research team members trained in human-subjects research and data confidentiality. The combined three-country dataset analyzed here was assembled by harmonizing the three site files into a single de-identified workbook; only de-identified, aggregate results are reported. In accordance with the approved protocol, electronic files will be retained for a minimum of three years following the study closure and permanently deleted thereafter; no paper records or biospecimens were collected.

2.6. Data Preparation and Variable Construction

Data preparation and statistical analyses were conducted in Python 3.12.3 using pandas 3.0.2, NumPy 2.4.4, SciPy 1.17.1, and statsmodels 0.14.6, with pairwise complete cases; no imputation was used. [Please remove: Data preparation was conducted in Python 3 using the pandas, NumPy, SciPy, and statsmodels libraries]. The following standardization steps were applied to the combined file before analysis:

Whitespace, stray non-breaking characters, and trailing-space variants in categorical fields (e.g., “Yes” vs. “Yes” in Prior experience; two encoding variants of the “>50” age category) were harmonized.

Familiarity (1–5), frequency (0 = Never to 4 = Daily), usefulness (1–5), and age group (1 = 18–21 to 7 = >50) were assigned ordinal numeric codes for non-parametric and regression analyses.

A composite perceived-impact score was computed as the unweighted mean of the seven impact items for each respondent; internal consistency was evaluated using Cronbach's α overall and within each site.

Discipline was derived from the open-ended course-enrolment item by keyword classification into broad categories (Health & Life Sciences; Social Sciences & Humanities; Natural Sciences & Engineering; Multiple-discipline; Other). This derived variable was used both as a stratifier in the analyses described in Sections 2.7.4 and 2.7.7 and as the basis of the sensitivity analysis in Section 2.7.9.

Tool prevalence was derived from the multiple-response “AI tools” and free-text “Most frequently used” fields by case-insensitive keyword matching against a pre-specified list of platforms.

The multiple-response items “Types of tasks” (eight task categories) and “What would encourage you to use AI tools more frequently” (five pre-specified factors) were each decomposed into a set of binary indicators by case-insensitive keyword matching of each respondent's response string against the pre-specified option labels. Each binary indicator was used as either an outcome (cross-site prevalence) or a predictor (selection vs. non-selection) in the analyses described in Section 2.7.7.

The seven perceived-impact items were stored in two label variants in the raw export (e.g., “(3) Moderate impact” and “Moderate impact”); both label variants were mapped to the same numeric code (1 = Not at all impact, 2 = Slight impact, 3 = Moderate impact, 4 = Strong impact, 5 = Significant impact).

Item-level missingness was concentrated in the impact, usefulness, frequency, task-type, and elaboration items, which were administered conditionally on AI adoption; missingness on these items is therefore structurally non-random and follows from the survey’s branching logic rather than from item-level non-response. Inferential analyses involving the impact items were restricted to confirmed AI users ($n = 409$) unless otherwise stated.

2.7. Statistical Analysis

All analyses were conducted in Python 3 (pandas 3.0.2, SciPy 1.17.1, and statsmodels 0.14.6). Generative AI tools (Claude, Anthropic, Opus 4.7) were used to assist with code drafting, data-visualization support, and language editing under direct author supervision; all statistical analyses, interpretations, and manuscript content were verified by the authors. Statistical significance was set at $\alpha = 0.05$ (two-tailed). Where multiple pairwise comparisons followed an omnibus test, p -values were Bonferroni-adjusted within the family of comparisons.

2.7.1. Hierarchy of Analyses: Primary, Secondary, and Exploratory

To guard against inflated Type I error across the many comparisons reported below, analyses are organized into three tiers and interpreted accordingly. The primary, confirmatory analyses are: (i) the cross-site comparison of AI adoption status (chi-square test of independence) and of the composite perceived-impact score among adopters (Kruskal–Wallis with Bonferroni-corrected pairwise contrasts); and (ii) the multivariable logistic regression of active AI use on site, gender, age group, prior CS/AI exposure, and AI familiarity. Secondary analyses include a multivariable OLS model of composite impact among adopters, a cross-site comparison of familiarity, frequency, and usefulness, and a sensitivity analysis restricted to Health & Life Sciences respondents. All other analyses—per-item impact comparisons across the seven impact domains, site-stratified correlations, multi-response task and encouragement-factor analyses, discipline-stratified comparisons, modality–tool and modality–task tests, and the open-ended thematic analysis—are treated as exploratory and reported descriptively. Exploratory tests are not formally adjusted beyond the corrections noted in the primary and secondary analyses, and their p -values should be read as effect-size descriptors rather than as confirmatory significance tests. The pre-registered VARK-related hypotheses (H1, H2) are also reported as exploratory because the dominance of text-based chatbots in the realized tool ecosystem (Section 3.10) effectively neutralizes the modality contrast presupposed by the hypotheses.

2.7.2. Descriptive Statistics

Categorical variables were summarized as frequencies and percentages; ordinal and continuous variables were summarized as means with standard deviations and medians. Composite-impact scores were reported overall and stratified by site.

2.7.3. Reliability

Internal consistency of the seven-item perceived-impact scale was evaluated using Cronbach’s α , computed on cases with complete data on all seven items, both overall and within each site sub-sample.

2.7.4. Bivariate Associations

Associations between AI adoption status (Yes/Maybe/No) and categorical predictors (site, gender, age group, first language, prior CS/AI exposure, familiarity) were evaluated using Pearson's chi-square test of independence with Cramér's V as an effect-size measure (interpretive guidelines: <0.10 negligible; 0.10–0.20 small; 0.20–0.40 moderate; >0.40 large). Tests were conducted on the pooled sample and stratified by site.

Cross-site comparisons of ordinal outcomes (familiarity, frequency, usefulness, composite impact, and the seven impact items) were conducted using the Kruskal–Wallis H test, followed by pairwise Mann–Whitney U tests with Bonferroni correction. Rank-biserial correlation was reported as a non-parametric effect-size estimate for pairwise contrasts.

2.7.5. Correlations

Associations among the ordinal variables of interest (familiarity, frequency, usefulness, composite impact, age) were estimated using Spearman's rank correlation coefficient, overall and stratified by site, and the sample was restricted to confirmed AI users.

2.7.6. Multivariable Models

To estimate the independent contribution of country and individual-level covariates to AI adoption, a multivariable binary logistic regression was fit with active AI use (Yes vs. No or Maybe) as the outcome and site, gender, age group (treated as ordinal), prior CS/AI exposure, and familiarity (treated as ordinal) as predictors. Adjusted odds ratios with 95% confidence intervals and McFadden's pseudo- R^2 are reported.

To estimate the determinants of perceived impact among adopters, an ordinary least-squares linear regression fit on the AI-user sub-samples ($n = 409$) with composite impact as the outcome and site, gender, age group, familiarity, and frequency as predictors. Unstandardized coefficients with 95% confidence intervals and the model R^2 are reported.

2.7.7. Multi-Response Item and Discipline Analyses

For each binary indicator derived from the multi-response items (eight task-category indicators from "Types of tasks"; five factor indicators from "What would encourage greater AI use"; the eight AI tool indicators described in Section 2.7), three families of analyses were conducted. Site differences in indicator prevalence were assessed using Pearson's chi-square test of independence with Cramér's V as an effect-size measure. For each task-category indicator, the difference in composite perceived impact between selectors and non-selectors was tested using the Mann–Whitney U test, with rank-biserial correlation reported as an effect-size estimate; the association between task breadth (the count of task categories selected per AI user, range 0–8) and composite impact was estimated using Spearman's rank correlation, overall and stratified by site. An exploratory per-item analysis of the seven impact items by task-category selection was also conducted (Mann–Whitney U) and is reported as exploratory.

The broad-discipline classification derived in Section 2.7 was treated as a five-level nominal stratifier. The association between discipline and AI adoption status was tested using Pearson's chi-square with Cramér's V. Cross-discipline differences in composite perceived impact among AI users were tested using the Kruskal–Wallis H test. Cross-discipline differences in familiarity and frequency were assessed analogously.

2.7.8. Exploratory Learning-Style Analyses

Consistent with the analytic hierarchy stated in Section 2.7.1, all learning-style analyses are treated as exploratory and hypothesis-generating. Three descriptive analyses were conducted. First, cross-site differences in modality endorsement (V, A, R, K) and in the

proportion classified as multimodal were summarized using Pearson's chi-square with Cramér's V. Second, the association between each modality-endorsement indicator and composite perceived impact among AI users was examined using the Mann–Whitney U test. Third, the IRB-protocol pre-registered hypotheses (H1: an association between VARK modality and category of AI tool adopted; H2: visual learners favor image-generation tools and reading/writing learners favor text-based chatbots) were evaluated descriptively via chi-square tests of independence between each modality indicator and the relevant tool-prevalence indicators (Section 2.7), and between modality endorsement and task-category selection. Because the realized tool ecosystem is dominated by text-based chatbots (Section 3.7), these tests were anticipated to have limited power for detecting modality-driven differentiation; the resulting *p*-values are reported as effect-size descriptors without further multiple-comparison adjustment, and any positive findings are flagged for replication.

2.7.9. Sensitivity Analysis

Because disciplinary composition differed across site sub-samples, a sensitivity analysis was conducted on the subset of respondents enrolled in Health & Life Sciences programs ($n = 272$), defined by keyword classification of the open-ended discipline field (programs containing references to medicine, microbiology, pharmacy, nursing, public health, biology, or life sciences). All primary site contrasts (adoption, familiarity, composite impact, usefulness) were re-estimated in this comparable-discipline subset to evaluate whether the site effects were robust to disciplinary mix.

2.7.10. Qualitative Analysis

Open-ended elaborations ($n = 66$ informative responses, U.S. sub-sample) were analyzed using an inductive–deductive hybrid thematic approach. Deductive scaffolding was provided by the survey's task-type categories; inductive coding identified emergent themes through iterative clustering. Themes were further mapped to Anderson and Krathwohl's revised Bloom's Taxonomy of cognitive processes (Remember, Understand, Apply, Analyze, Evaluate, Create) to characterize the cognitive level of reported AI use. Quantitative thematic and Bloom's-level analysis of the Indonesian and Tajik open-ended responses was not undertaken in the present study and is identified as a priority for future work.

3. Results

3.1. Sample Characteristics

The combined analytic dataset comprised 584 valid undergraduate responses across three sites: Indonesia ($n = 235$; 40.2%), Tajikistan ($n = 226$; 38.7%), and the United States ($n = 123$; 21.1%). Most respondents were aged 18–21 years ($n = 380$; 65.1%), with 22.4% aged 22–25 and the remainder distributed across older age categories. Female respondents constituted 56.5% of the combined sample ($n = 330$), male 42.6% ($n = 249$), and 0.7% preferred not to disclose. Across the full sample, 47.4% ($n = 276$) reported prior formal exposure to computer science, data science, or AI-related fields. Internal consistency of the seven-item perceived-impact scale was excellent in the pooled sample (Cronbach's $\alpha = 0.907$; $n = 409$ complete cases) and within each site sub-sample (United States $\alpha = 0.922$, $n = 58$; Tajikistan $\alpha = 0.882$, $n = 153$; Indonesia $\alpha = 0.919$, $n = 198$), justifying treatment of the unweighted mean of the seven items as a single composite-impact score.

3.2. AI Adoption by Site

AI adoption differed substantially across the three sites (Figure 1). Overall, 70.0% of respondents reported active AI use, 14.7% responded “Maybe”, and 15.2% reported

non-use. The adoption gradient was Indonesia (84.3%) > Tajikistan (67.7%) > United States (47.2%); the omnibus association between country and adoption status was statistically significant and of moderate magnitude ($\chi^2(4) = 83.22, p < 0.001$, Cramér's $V = 0.267$).

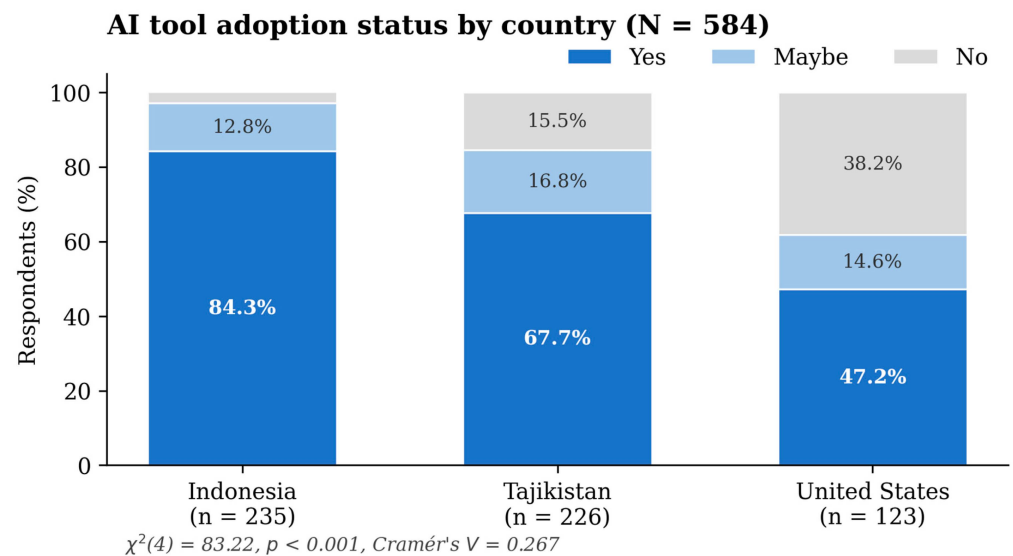


Figure 1. AI tool adoption status by country (N = 584). Stacked bars show the percentage of respondents reporting active AI use (“Yes”), prospective use (“Maybe”), and non-use (“No”). The adoption gradient is Indonesia > Tajikistan > United States; the omnibus test of independence is statistically significant ($\chi^2(4) = 83.22, p < 0.001$, Cramér's $V = 0.267$, moderate effect).

3.3. Familiarity, Frequency, and Usefulness

Cross-site differences were observed on all three engagement indicators (Figure 2). Tajik respondents reported the highest self-rated familiarity ($M = 3.59$, $SD = 1.19$) and the highest frequency of use among AI users and prospective users ($M = 3.08$, $SD = 1.08$). U.S. respondents reported the lowest familiarity ($M = 2.98$, $SD = 1.11$) and the lowest frequency ($M = 2.34$, $SD = 1.21$) but, among adopters, the highest mean overall usefulness rating ($M = 4.03$, $SD = 0.86$).

Familiarity, frequency, and perceived usefulness by country

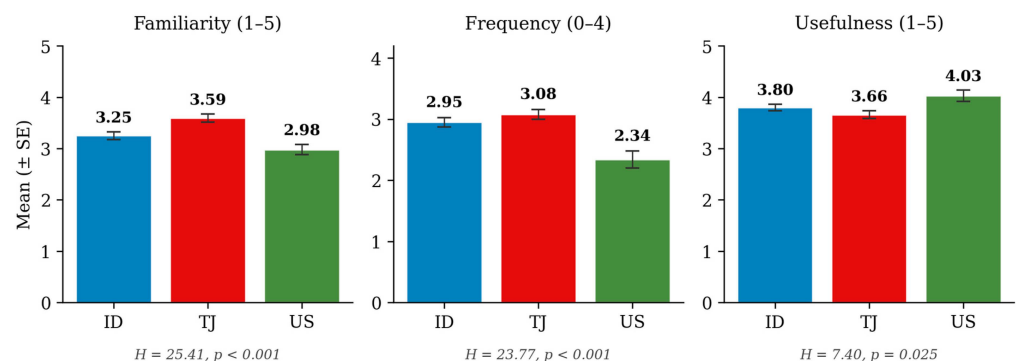


Figure 2. Cross-site comparisons of self-rated familiarity (1–5), frequency of use (0–4), and overall perceived usefulness (1–5). Bars show site means with standard error bars. Kruskal–Wallis omnibus statistics are shown beneath each panel.

3.4. Perceived Academic Impact

Indonesian adopters reported the highest overall perceived impact, with Tajikistan trailing as the only statistically distinguishable contrast. Mean composite scores were 3.27

(SD = 0.79) in Indonesia, 3.07 (SD = 0.97) in the United States, and 2.96 (SD = 0.80) in Tajikistan (Kruskal–Wallis $H = 15.34$, $p < 0.001$). Bonferroni-corrected Mann–Whitney pairwise contrasts identified only the Tajikistan–Indonesia difference as significant ($U = 11,446$, $p = 0.0002$, rank-biserial $r = 0.244$). Because formal measurement invariance of the composite scale across sites was not tested, these cross-site mean comparisons should be read as provisional [remove the reference (see Section 4)].

The cross-site variation in perceived impact is concentrated in writing quality, data analysis, and collaboration; foundational tasks such as learning new material and assignment efficiency are perceived similarly across all three settings (Table 1; Figure 3). Despite these domain-specific differences, the within-site rank ordering of impact domains is highly conserved (Spearman ρ between-site profiles: U.S.–TJ 0.82; U.S.–ID 0.64; TJ–ID 0.68), indicating that students at all three sites agree on which kinds of tasks AI helps most, even where they disagree on the magnitude of that help.

Table 1. Mean per-item perceived impact (1–5 scale) by site, AI users only.

Domain	US M (SD)	TJ M (SD)	ID M (SD)	H	<i>p</i>
Learning new material	3.35 (1.10)	3.24 (1.09)	3.37 (0.93)	2.36	0.307
Efficiency	3.22 (1.16)	3.18 (0.97)	3.33 (0.99)	3.30	0.192
Writing quality	3.19 (1.07)	2.80 (1.06)	3.30 (0.99)	23.95	<0.001
Creativity	2.91 (1.20)	2.95 (1.05)	3.26 (0.99)	11.36	0.003
Exam preparation	3.31 (1.19)	2.99 (1.13)	3.22 (0.95)	6.09	0.048
Data analysis	3.05 (1.19)	2.82 (0.97)	3.18 (0.89)	15.56	<0.001
Collaboration	2.45 (1.27)	2.72 (1.05)	3.20 (0.97)	31.35	<0.001

Kruskal–Wallis omnibus tests. Sample sizes: US $n = 58$, TJ $n = 153$, ID $n = 198$.

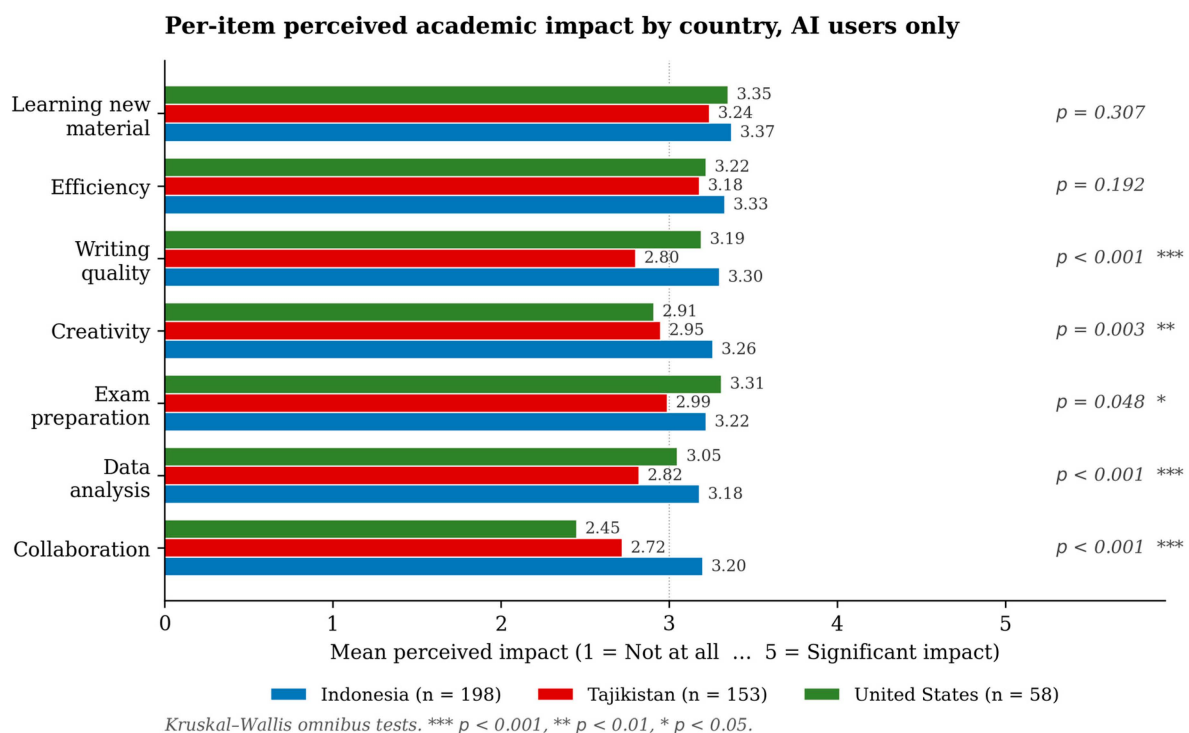


Figure 3. Per-item perceived academic impact (1–5 scale) by site, AI users only. Bars show site means for each of the seven impact domains; the dotted vertical line marks the scale midpoint (3 = Moderate impact). Kruskal–Wallis omnibus p -values are shown to the right of each domain.

In the pooled sample, AI adoption status was significantly associated with site ($V = 0.267$), first language ($V = 0.269$), familiarity ($V = 0.255$), prior CS/AI exposure ($V = 0.234$), and age group ($V = 0.196$); gender was not significantly associated with adop-

tion ($V = 0.075$, $p = 0.159$). Stratified analyses showed that familiarity was the only predictor significantly associated with adoption at all three sites (United States $V = 0.307$, $p = 0.003$; Tajikistan $V = 0.318$, $p < 0.001$; Indonesia $V = 0.261$, $p < 0.001$).

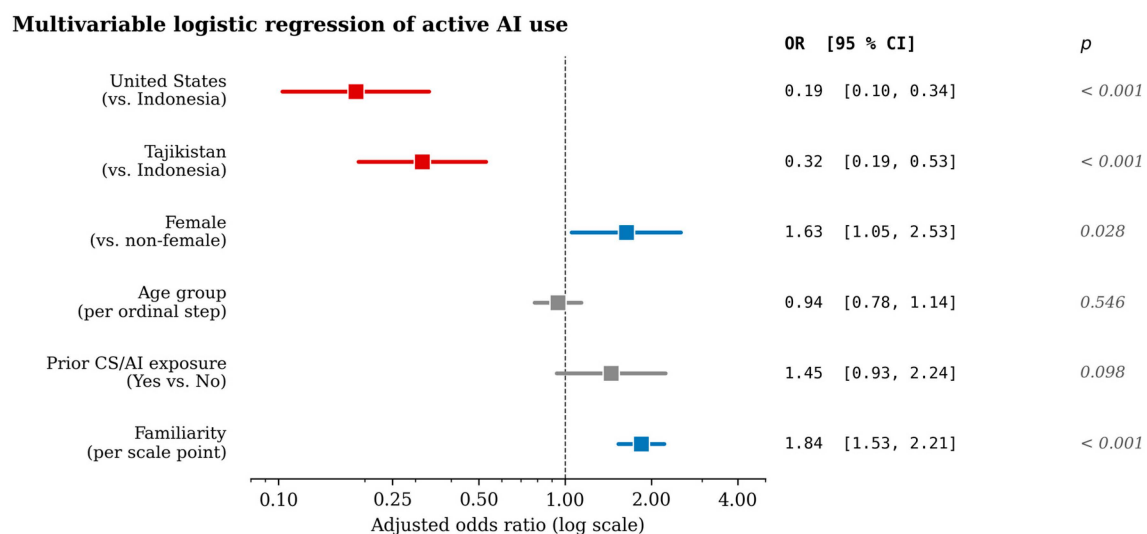
3.5. Multivariable Logistic Regression of AI Adoption

A multivariable logistic regression of active AI use on site, gender, age group, prior CS/AI exposure, and familiarity ($n = 584$; McFadden's pseudo- $R^2 = 0.164$) confirmed that site effects persisted after adjustment (Table 2; Figure 4). U.S. respondents had odds of active AI use 81% lower than Indonesian respondents (adjusted OR = 0.19, 95% CI: 0.10–0.33), and Tajik respondents had odds 68% lower (adjusted OR = 0.32, 95% CI: 0.19–0.53). Each one-point increase in familiarity was associated with 84% higher odds of active use (adjusted OR = 1.84, 95% CI: 1.53–2.22). Female respondents were more likely than non-female respondents to be active users after adjustment (adjusted OR = 1.63, $p = 0.028$).

Table 2. Multivariable logistic regression of active AI use ($n = 584$).

Predictor	Adjusted OR	95% CI	<i>p</i>
United States (vs. Indonesia)	0.19	0.10–0.33	<0.001
Tajikistan (vs. Indonesia)	0.32	0.19–0.53	<0.001
Female (vs. non-female)	1.63	1.05–2.53	0.028
Age group (per ordinal step)	0.94	0.78–1.14	0.546
Prior CS/AI exposure (Yes vs. No)	1.45	0.93–2.24	0.098
Familiarity (per scale point)	1.84	1.53–2.22	<0.001

Reference categories: Country = Indonesia. Gender = non-female. Prior CS exposure = No. McFadden's pseudo- $R^2 = 0.164$.



$n = 583$; McFadden's pseudo- $R^2 = 0.164$. Reference categories: Country = Indonesia; Gender = non-female; Prior CS = No.

Figure 4. Forest plot of adjusted odds ratios from the multivariable logistic regression ($n = 584$). Red: significant protective effect ($OR < 1$, $p < 0.05$); blue: significant positive effect ($OR > 1$, $p < 0.05$); gray: non-significant.

3.6. Correlates of Perceived Impact

Among AI users, the composite perceived-impact score was moderately correlated with familiarity (Spearman $\rho = 0.371$, $p < 0.001$), frequency ($\rho = 0.325$, $p < 0.001$), and usefulness ($\rho = 0.327$, $p < 0.001$). Country-stratified correlations revealed notable heterogeneity in the familiarity–frequency association (U.S., $\rho = 0.584$; Tajikistan, $\rho = 0.208$; Indonesia, $\rho = 0.119$), whereas the familiarity–impact correlation was moderate and significant in all countries (Figure 5).

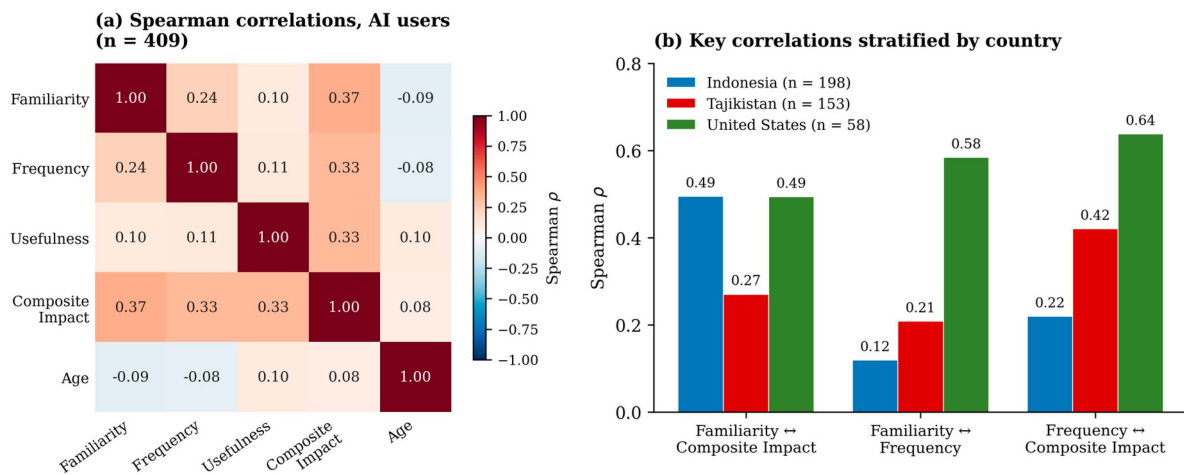


Figure 5. Spearman correlations. (a) Correlation matrix among AI users ($n = 409$). (b) Key bivariate correlations stratified by site.

A multivariable OLS regression of composite impact among AI users ($n = 409$; $R^2 = 0.269$) identified familiarity ($\beta = 0.270$, $p < 0.001$) and frequency ($\beta = 0.206$, $p < 0.001$) as independent positive predictors of impact. After adjustment, Tajik respondents reported lower composite impact than Indonesian respondents ($\beta = -0.483$, $p < 0.001$); the U.S.–Indonesia gap was not significant ($\beta = -0.158$, $p = 0.168$).

3.7. Tool Ecosystem

Three distinct site-level tool ecosystems were observed (Table 3). ChatGPT was dominant at every site; Grammarly was substantially more prevalent in the United States (51.3%) than at the other two sites; Perplexity AI showed a distinctive Indonesian signature (46.1% vs. 3.7% and 2.6%); QuillBot was endorsed by 24.6% of Indonesian respondents but was effectively absent in Tajikistan.

Table 3. AI tool prevalence among AI users and prospective users (Yes + Maybe), percent.

Tool	Indonesia (n = 228)	Tajikistan (n = 191)	United States (n = 76)
ChatGPT	89.5	96.3	72.4
Gemini	67.5	69.1	31.6
Perplexity AI	46.1	3.7	2.6
Grammarly	31.6	2.6	51.3
QuillBot	24.6	0.0	14.5
Microsoft Copilot	12.7	7.3	13.2
DeepSeek	1.8	3.7	1.3
NotebookLM	0.0	1.0	2.6

Multiple selections permitted; derived by keyword matching of the multi-response and free-text items.

3.8. Effect of AI on Specific Learning Tasks

Indonesian and U.S. adopters use AI for a wider range of tasks than Tajik adopters. The mean number of task categories selected per AI user ($n = 409$) was 2.78 in Indonesia, 2.76 in the United States, and 2.33 in Tajikistan (overall $M = 2.62$, $SD = 1.66$; Kruskal–Wallis $H = 12.63$, $p = 0.002$). The eight categories examined were studying and learning, brainstorming, writing assistance, coding/programming, data analysis or visualization, language learning, presentation and design, and administrative tasks. These task analyses are reported as exploratory.

Studying and learning are the dominant use cases across all sites (79.5% overall), but secondary use cases vary sharply. The largest cross-site contrast is for brainstorming: U.S.

(58.6%) and Indonesian (54.5%) adopters use AI for brainstorming at more than twice the rate of Tajik adopters (24.8%; $\chi^2 = 36.77$, $V = 0.30$, the largest task-level effect observed). Writing assistance follows a similar gradient (U.S. 55.2%, Indonesia 46.0%, Tajikistan 32.0%; $V = 0.17$). Two sites depart from this pattern in informative ways: administrative use is highest in the United States (24.1% vs. 8.6% in Indonesia; $V = 0.16$), and coding use is highest—though still modest—in Tajikistan (12.4% vs. 3.5% U.S.; Figure 6a, Table 4).

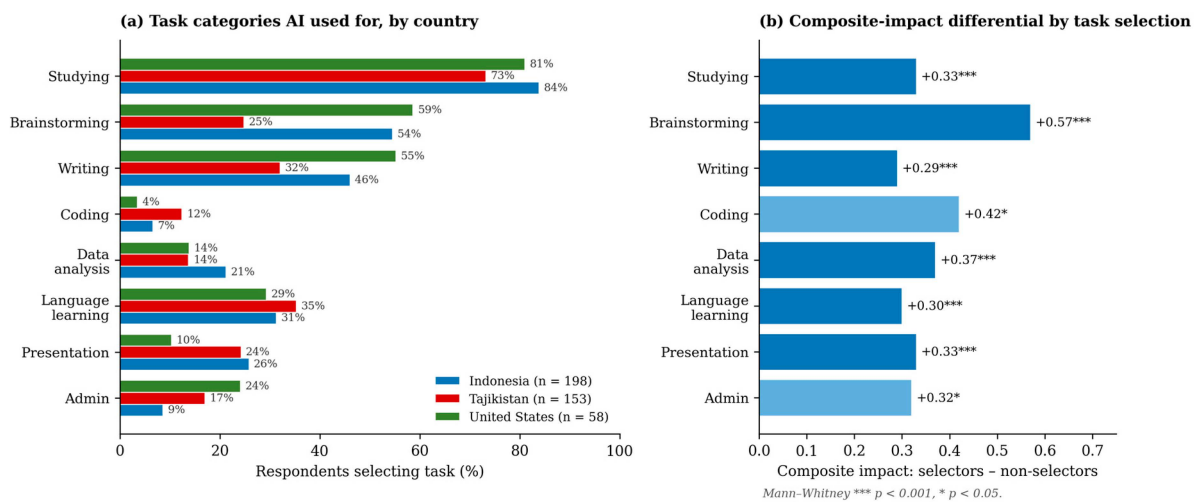


Figure 6. Task usage patterns and perceived-impact differentials. (a) Percentage of AI users selecting each of the eight task categories, stratified by site. Brainstorming, Writing assistance, Coding, Presentation, and Administrative tasks showed significant site differences. (b) Mean composite-impact differential (selectors minus non-selectors) for each task category. All eight categories were associated with significantly higher perceived impact among selectors.

Table 4. Task-category selection (%) by country among AI users ($n = 409$).

Task Category	ID	TJ	US	Overall	χ^2	p	Cramér's V
Studying and learning	83.8	73.2	81.0	79.5	6.08	0.05	0.12
Brainstorming ideas	54.5	24.8	58.6	44.0	36.77	<0.001	0.30
Writing assistance	46.0	32.0	55.2	42.1	11.65	0.003	0.17
Coding/programming	6.6	12.4	3.5	8.3	5.98	0.05	0.12
Data analysis/visualization	21.2	13.7	13.8	17.4	3.97	0.137	0.10
Language learning	31.3	35.3	29.3	32.5	0.94	0.624	0.05
Presentation/design	25.8	24.2	10.3	23.0	6.22	0.045	0.12
Administrative tasks	8.6	17.0	24.1	13.9	10.95	0.004	0.16

Chi-square tests of independence with Cramér's V. Site sample sizes: Indonesia $n = 198$, Tajikistan $n = 153$, United States $n = 58$.

Brainstorming emerged as the task most strongly associated with perceived academic benefit. Adopters who used AI for brainstorming reported substantially higher composite impact than non-selectors ($M = 3.44$ vs. 2.87 ; rank-biserial $r = 0.41$, $p < 0.001$, Figure 6b), an effect roughly twice the size of the next-largest differentials (coding $r = 0.26$; data analysis $r = 0.25$; studying and learning $r = 0.23$). All eight task categories showed positive selector–non-selector differentials (all $p < 0.05$), and task breadth—the number of categories endorsed—correlated moderately with composite impact at every site (Spearman $\rho = 0.34$ – 0.47 , all $p < 0.001$). These selector–non-selector differentials and breadth correlations are reported here descriptively; they are cross-sectional and exploratory, and their interpretation is developed in Section 4.

An exploratory per-item analysis (not tabled) showed intuitive content alignment: writing-assistance selectors reported the largest differential on the writing-quality item,

data-analysis selectors on the data-analysis item, and brainstorming selectors on every item. The pattern suggests that respondents attribute benefit specifically to the task for which they use AI; this domain-by-task analysis is reported descriptively.

3.9. Factors That Would Encourage Greater AI Use

The multiple-response item “What would encourage you to use AI tools more frequently in your studies” presented five pre-specified options. The item was administered to all AI users and prospective users (Yes or Maybe, $n = 495$); respondents who reported no AI use did not see this item. Overall, the most endorsed factors were more user-friendly tools (45.2%) and Formal training on AI tools (43.8%), followed by curriculum integration (20.2%), faculty endorsement (19.5%), and Demonstrated success (17.5%).

Site differences were observed on three of the five factors (Table 5). Tajik respondents endorsed more user-friendly tools at a substantially higher rate than respondents at the other two sites (58.4% vs. Indonesia 45.1% vs. United States 21.1%; $V = 0.28$). Curriculum integration was endorsed by 39.2% of Indonesian respondents but by only 0.9% of Tajik respondents ($V = 0.42$, the largest cross-site effect observed in the dataset). Formal training endorsement was lower in the United States (29.3%) than in Indonesia (48.5%) or Tajikistan (46.9%; $V = 0.15$). Faculty endorsement and Demonstrated success did not show significant site differences.

Table 5. Endorsement (%) of factors that would encourage greater AI use, by site (AI users + prospective users, $n = 495$).

Factor	ID ($n = 228$)	TJ ($n = 191$)	US ($n = 76$)	Overall	χ^2	p	Cramér's V
Formal training on AI tools	48.5	46.9	29.3	43.8	13.55	<0.001	0.15
More user-friendly AI tools	45.1	58.4	21.1	45.2	44.67	<0.001	0.28
Curriculum integration	39.2	0.9	19.5	20.2	104.67	<0.001	0.42
Faculty endorsement	17.0	18.6	26.0	19.5	4.36	0.113	0.09
Demonstrated success of AI	19.6	19.0	10.6	17.5	5.17	0.075	0.09

Multiple selections permitted. The curriculum-integration contrast ($V = 0.42$) is the largest cross-site effect observed in the study.

Two cross-tabulations are reported. First, more user-friendly tools were endorsed at 54.3% among confirmed users and 48.8% among prospective “Maybe” users. Second, Formal training was the most-endorsed factor among prospective users in Tajikistan (47.7%) and the second-most-endorsed in Indonesia (47.7%). In the United States, faculty endorsement (26.0%) was the most-endorsed factor—the only site at which faculty endorsement reached the top tier. These patterns are interpreted in Section 4.

3.10. Effect of Academic Discipline

The open-ended discipline field was classified into five broad categories (Health & Life Sciences; Social Sciences & Humanities; Natural Sciences & Engineering; Multiple-discipline; Other), based on keyword matching of the enrolled program. Of the 584 respondents, 524 (89.7%) could be classified into one of these categories; 60 (10.3%) had blank or uncategorizable entries and were excluded from discipline-stratified analyses. Discipline composition differed sharply across sites (Figure 7c): Tajikistan was overwhelmingly Health & Life Sciences (83.6% of respondents), Indonesia was concentrated in Social Sciences/Humanities (41.3%) and Other programs (28.9%), and the United States combined Multiple-discipline (43.9%), Health & Life Sciences (35.8%), and Natural Sciences/Engineering (13.8%) enrolments.

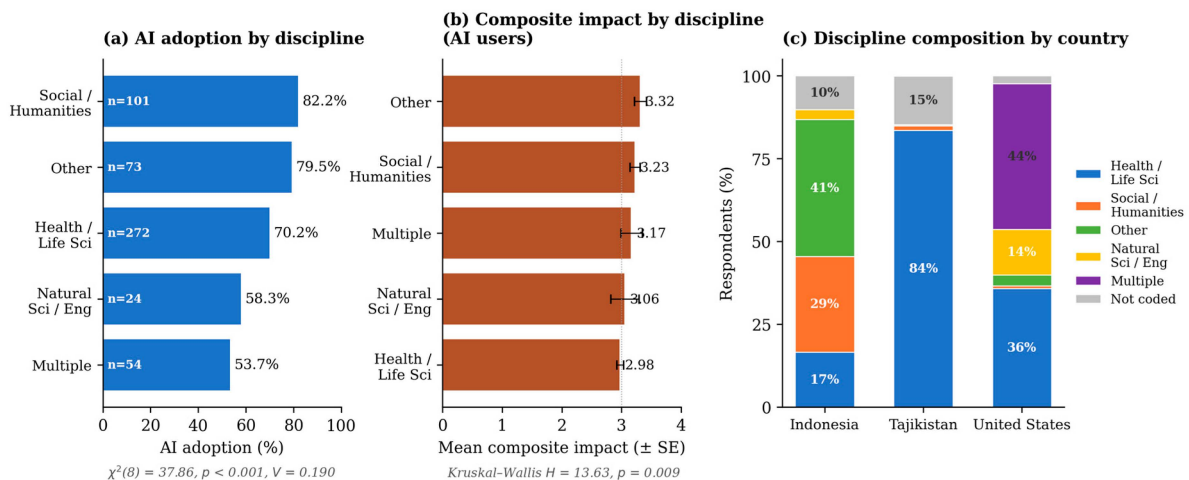


Figure 7. Effect of academic discipline. (a) AI adoption (% Yes) varies significantly across the five broad disciplines ($\chi^2(8) = 37.86, p < 0.001$). (b) Mean composite perceived impact among AI users by discipline; the Health & Life Sciences subgroup, the largest in the dataset, reports the lowest impact (Kruskal-Wallis $H = 13.63, p = 0.009$). (c) Discipline composition by site, showing the predominantly Health & Life Sciences enrolment in Tajikistan, the diversified mix in Indonesia, and the Multiple-discipline/Health & Life Sciences mix in the United States.

AI adoption varied significantly across disciplines ($\chi^2(8) = 37.86, p < 0.001$, Cramér's $V = 0.190$; Figure 7a). The highest adoption rates were observed in Social Sciences & Humanities (82.2%) and Other programs (79.5%), followed by Health & Life Sciences (70.2%), Natural Sciences & Engineering (58.3%), and Multiple-discipline enrollments (53.7%). The Multiple-discipline category—composed almost entirely of U.S. respondents who indicated more than one program—had the lowest adoption rate of any group, consistent with the overall U.S. adoption pattern reported in Section 3.2.

Composite perceived impact also differed across disciplines (Kruskal-Wallis $H = 13.63, p = 0.009$; Figure 7b). Among AI users, Other programs ($M = 3.32, n = 58$) and Social Sciences & Humanities ($M = 3.23, n = 83$) reported the highest impact, followed by Multiple-discipline ($M = 3.17, n = 29$) and Natural Sciences & Engineering ($M = 3.06, n = 14$); Health & Life Sciences ($M = 2.98, n = 191$) reported the lowest impact. The Health & Life Sciences result is notable because this is by far the largest disciplinary subgroup and is the principal driver of the relatively low Tajik composite-impact mean reported in Section 3.4; the two findings are mutually reinforcing.

Cross-site sensitivity testing on the Health & Life Sciences subset demonstrated that the principal cross-site contrasts in adoption, familiarity, and composite impact are preserved when disciplinary mix is held approximately constant, indicating that the site effects reported in Sections 3.2–3.4 are not artifacts of differential disciplinary composition. The disciplinary pattern documented here is therefore best interpreted as an independent layer of variation that operates alongside, rather than in place of, the site effects.

3.11. Learning-Style Preferences and AI Engagement (Exploratory)

Exploratory learning-style analyses. Exploratory analyses of VARK learning-style preferences found little evidence that modality preferences meaningfully influenced generative AI adoption or perceived academic impact. Most respondents reported multimodal learning preferences, and although modest differences in modality endorsement were observed across study sites, only a small association between Reading/Writing preference and perceived impact reached nominal significance. The pre-registered hypotheses linking learning-style preferences to specific AI tool categories could not be adequately evaluated

because AI use was overwhelmingly dominated by text-based chatbot platforms across all three sites. Detailed analyses and tables are provided in the Supplementary Materials; the interpretation of these exploratory findings is developed in Section 4.

4. Discussion

This study examined AI tool adoption and perceived academic impact among 584 undergraduates across three institutions in Indonesia, Tajikistan, and the United States. Four findings stand out. First, AI adoption varied substantially across the three sites, with a pronounced gradient (Indonesia > Tajikistan > United States) that persisted after multivariable adjustment for individual-level covariates. Second, self-rated familiarity with AI was the most generalizable individual-level correlate of both adoption and perceived impact across all three settings. Third, the seven-item perceived-impact scale showed excellent internal consistency overall and within each site (Cronbach's $\alpha \geq 0.88$); however, cross-site mean comparisons remain provisional pending formal measurement-invariance testing (see Section 4). Fourth, three distinct site-level tool ecosystems emerged—a writing-tool-heavy U.S. profile, a chatbot-only Tajik profile, and a diverse Indonesian profile in which Perplexity AI achieved unusually high penetration.

The cross-site pattern is not a single “more is better” gradient. Indonesia combined the broadest adoption (84.3%) with the highest perceived impact among adopters, the United States combined the narrowest adoption (47.2%) with the highest usefulness rating among its smaller adopter pool, and Tajikistan combined broad adoption and the highest frequency of use with the lowest composite impact. These three configurations suggest that perceived value is contingent on contextual factors not captured by individual-level variables alone—most plausibly, institutional policy, faculty endorsement, the language of instruction, and the availability of task-specific tools.

The multivariable logistic regression supports this contextual interpretation. After adjusting for gender, age, prior computing exposure, and AI familiarity, the site effect remained large (adjusted ORs of 0.19 for the U.S. and 0.32 for Tajikistan, relative to Indonesia), indicating that site captures unmeasured contextual differences operating above and beyond individual covariates.

Self-rated familiarity was the only individual-level variable to operate consistently across all three sites, predicting both adoption (Cramér's $V = 0.26\text{--}0.32$) and perceived impact (Spearman $\rho = 0.27\text{--}0.50$) at every site, and emerging as a strong independent predictor in both multivariable models. This consistency aligns with technology acceptance theory (Sergeeva et al., 2025), in which perceived ease of use and self-efficacy are stable proximal antecedents of adoption. This result should not, however, be over-read. Self-rated familiarity is a subjective indicator of confidence and prior exposure; it is a proxy for, and not a measurement of, AI literacy as defined in our conceptual framework—a multidimensional construct spanning conceptual understanding, evaluative judgment, and ethical awareness (Chiu, 2025; Ng et al., 2021). Our findings therefore support interventions that raise students' confidence with, and exposure to, generative AI tools, and suggest that such interventions may translate more uniformly across institutional contexts than interventions targeting other levers. Whether interventions targeting AI literacy proper yield comparable gains in adoption or perceived impact cannot be established from these data and would require direct measurement of the broader construct.

Country-stratified correlations also revealed a meaningful heterogeneity in the familiarity–frequency relationship: strong in the U.S. ($\rho = 0.58$), moderate in Tajikistan ($\rho = 0.21$), and weak in Indonesia ($\rho = 0.12$). One interpretation is that in higher-adoption settings, frequent use has become the default, no longer distinguishing self-rated experts from novices, whereas in lower-adoption settings, familiarity continues to closely track

usage. If this interpretation is correct, self-rated familiarity will become a progressively less informative discriminator of use as normalization advances.

All eight task categories were associated with significantly higher perceived impact among selectors than non-selectors, with the largest differential for brainstorming ($\Delta = +0.57$, rank-biserial $r = 0.41$, $p < 0.001$). Per-item analyses showed intuitive content alignment: writing-task selectors reported the largest gains on the writing-quality item, data-analysis selectors on the data-analysis item, and brainstorming selectors on every item. Task breadth—the number of categories for which a respondent reported using AI—correlated moderately with composite impact at every site (Spearman $\rho = 0.34$ – 0.47). Taken together, these patterns suggest that students accurately attribute benefits to the tasks for which they actually use AI, and that diversified, fit-for-purpose use is a more promising target than intensified use of a single tool. This task-specific patterning is consistent with prior empirical work (Chan & Hu, 2023; Ravšelj et al., 2025). Our contribution is to show that, at all three sites, it is the breadth of the task repertoire—rather than the intensity of use of any single tool—that tracks perceived benefit.

Domain-specific site differences were concentrated in writing quality, data analysis, and collaboration, with the largest absolute gap for collaboration (Indonesia $M = 3.20$ vs. U.S. $M = 2.45$). The writing-quality gap maps onto the tool-ecology data: Tajik respondents reported the lowest writing-quality impact and used essentially no writing-specialized tools (Grammarly 2.6%; QuillBot 0%), whereas U.S. and Indonesian respondents reported higher writing-quality impact and used these tools at substantially higher rates. The cross-sectional design precludes causal inference, but tool diversity and task-tool matching are plausible mediators of the observed site differences in domain-specific impact.

The largest cross-site effect in the entire dataset was on the encouragement factor “curriculum integration of AI tools” (Cramér’s $V = 0.42$), with endorsement rates of 39.2% in Indonesia, 19.5% in the United States, and 0.9% in Tajikistan. The Tajik near-zero endorsement most plausibly reflects a context in which broad individual-level AI use has developed independently of any curriculum-level conversation. The Indonesian endorsement is consistent with the broad, diverse, high-impact adoption observed there. The U.S. profile is distinctive in a third way: faculty endorsement was the most-endorsed encouragement factor only in the United States (26.0%), consistent with a more individualized, faculty-led mode of adoption.

These three configurations argue strongly that institution-level interventions—clear AI policy, faculty endorsement, licensed tool availability, and language of instruction—are likely to outperform purely individual-targeted interventions of comparable cost. They also argue against a one-size-fits-all institutional template: the most effective intervention at one site (curriculum integration in Indonesia) is the least-endorsed factor at another (Tajikistan).

The encouragement-factor cross-tabulations reported in Section 3.9 support the same reading. That “more user-friendly tools” was endorsed at a comparable rate by confirmed users (54.3%) and by prospective “Maybe” users (48.8%) indicates that the perceived usability barrier is not dissolved by experience alone. That “formal training” was the most-endorsed factor among prospective users in Tajikistan (47.7%) and the second-most-endorsed in Indonesia (47.7%) identifies training provision as a candidate high-leverage intervention at those two sites—though, consistent with the distinction drawn above, what these students are asking for is exposure and confidence-building, not necessarily AI literacy in the fuller sense. The prominence of faculty endorsement in the United States (26.0%) is, by contrast, consistent with the more individualized, faculty-mediated mode of adoption visible in the U.S. familiarity–frequency correlation (Section 3.3). That the salient lever differs so sharply across three sites echoes the marked setting-to-setting variation in

student uptake and attitudes documented in large multi-institution surveys (Ravšelj et al., 2025) and reinforces the case against a single institutional template.

The pre-registered modality–tool matching hypotheses (H1: an association between VARK modality and AI tool category; H2: visual learners would favor image-generation tools and reading/writing learners would favor text-based chatbots) cannot be meaningfully tested in the present sample, so we treat them as exploratory rather than confirmed or rejected. The image-tool prediction collides with an extreme floor: image-generation use is <3% of AI users, leaving no power to detect modality differentiation. The chatbot prediction collides with an extreme ceiling: chatbot adoption is 95.8% across all users. In both cases, the realized ecosystem rather than the data itself defeats the test.

A small reading/writing–modality advantage in perceived impact did emerge ($M = 3.23$ vs. 3.08 , $p = 0.046$), consistent with the text-centric affordances of dominant tools, but the magnitude is too small to support strong claims. The broader pattern is that ecosystem-level standardization on chatbot interfaces leaves little room for learning-style differentiation in the current tool environment, independent of any theoretical position on the validity of the VARK framework itself (Pashler et al., 2008). Meaningful future tests will require either modality-differentiated tools or behavioral rather than self-report outcomes.

Two interpretive limitations warrant particular emphasis because they shape how the cross-site findings should be read. First, formal measurement invariance of the seven-item perceived-impact scale across sites was not tested. Although internal consistency was excellent within each sub-sample (Cronbach's $\alpha \geq 0.88$), cross-site mean comparisons of the composite or its individual items implicitly assume that respondents in Indonesia, Tajikistan, and the United States interpret the items and use the response scale equivalently—an assumption that translation, cultural response style, and discipline-specific framing can all violate. Without configural, metric, and scalar invariance testing via confirmatory factor analysis, the observed site differences in mean impact may reflect measurement non-equivalence rather than substantive differences in perceived benefit. We therefore treat the cross-site mean contrasts (notably the Indonesia–Tajikistan composite-impact gap and the domain-specific contrasts in Section 3.4) as provisional until invariance has been established. The language of administration also differed across sites (English in the United States and Indonesia, Russian in Tajikistan). Although Tajik respondents are fluent in Russian and the instrument underwent translation and independent back-translation by two bilingual authors with item content held constant, residual non-equivalence cannot be excluded; in particular, the Indonesian instrument, delivered in English, was completed largely by non-native English speakers. These features could produce item functioning that differs across sites independently of any true difference in perceived impact; established procedures for testing and reporting such invariance are reviewed elsewhere (Putnick & Bornstein, 2016; Vandenberg & Lance, 2000). Second, the manuscript reports a large number of statistical tests, including per-item, subgroup, modality, task-specific, and discipline-stratified analyses. While the hierarchy of analyses set out in Section 2.7 distinguishes confirmatory, secondary, and exploratory tests, the exploratory tier inflates the family-wise error rate and is best read as effect-size description rather than confirmatory hypothesis testing.

Beyond these two interpretive limitations, further methodological constraints should be noted. The cross-sectional self-report design precludes causal inference: observed associations between familiarity, frequency, and perceived impact are equally consistent with reverse causation or shared unmeasured determinants. The three sub-samples differed in disciplinary mix; although the sensitivity analysis indicated the principal cross-site contrasts are robust to disciplinary restriction, residual confounding by program type, year of study, or institutional culture cannot be excluded. The U.S. sub-sample ($n = 123$) was approximately half the size of either partner sample, reducing power for U.S.-specific

contrasts. Accordingly, non-significant results within the smaller U.S. sample and within other small subgroups should be read as inconclusive rather than as positive evidence of no effect. The frequency item used a four-level ordinal scale that is coarse for intensity-of-use questions. Open-ended elaborations were coded thematically only for the U.S. sub-sample; corresponding qualitative work for the Indonesian and Tajik sub-samples remains to be undertaken. Tool-prevalence estimates rely on keyword matching against free-text responses and may underestimate platforms referred to inconsistently or under alternative names.

The findings suggest two practical directions. Institutionally, the persistence of large site effects after adjustment for individual-level covariates points to institutional context—AI policy guidance, faculty endorsement, language of instruction, and campus-licensed tool availability—as a determinant of AI engagement; institution-level interventions are therefore likely to yield gains that purely individual interventions cannot. Pedagogically, the consistent familiarity–adoption–impact triad supports brief, scalable modules that build hands-on exposure to, and confidence with, generative AI tools—a candidate high-leverage intervention applicable across diverse settings. We frame this deliberately as a familiarity-building rather than an AI literacy recommendation: because AI literacy was not directly measured here, the additional value of instruction in critical evaluation, model limitations, and ethical use, although independently well motivated (Chiu, 2025; Ng et al., 2021), is not evidenced by the present findings.

Methodologically, the most urgent extension is confirmatory factor analysis with measurement-invariance testing across sites; this is a prerequisite, not an enhancement, for substantive cross-site mean comparisons of the perceived-impact composite. Three further extensions are warranted: (i) ordinal and multinomial models preserving the three-level adoption outcome (Yes/Maybe/No) to distinguish hesitant from refusing non-users; (ii) latent-class analysis of the joint distribution of frequency, familiarity, usefulness, and task profile to identify user typologies and test whether typology membership varies by site; (iii) future data collection that includes institutional-context variables (AI policy, faculty endorsement, licensed tool list), a brief AI literacy or critical-evaluation scale, and a longitudinal component permitting estimation of the link between perceived impact and measured academic outcomes.

5. Conclusions

Across 584 undergraduates at three institutions in Indonesia, Tajikistan, and the United States, AI tool adoption was widespread but site-patterned: Indonesia showed the broadest adoption and the highest perceived academic impact; the United States showed the narrowest adoption but the highest mean usefulness rating among adopters. These site differences were large and persisted after multivariable adjustment for individual-level covariates, indicating that unmeasured contextual factors continue to shape AI engagement. Self-rated familiarity emerged as the most generalizable individual-level correlate of adoption and perceived impact across all three settings, and three distinct site-level tool ecosystems were observed, with domain-specific impact differences mapping onto these ecosystems in interpretable ways—particularly for writing-related outcomes.

The seven-item perceived-impact scale showed excellent internal consistency in all three site sub-samples and provides a reusable instrument for cross-institutional comparison; formal measurement-invariance testing is a prerequisite for substantive interpretation of cross-site means. The findings support institution-level interventions that combine structured opportunities to build familiarity with generative AI tools with explicit policy guidance and faculty endorsement. Because familiarity was self-rated and AI literacy was not directly measured, claims about the efficacy of AI literacy interventions remain a

hypothesis for future work, alongside measurement invariance, user typologies, and the relationship between perceived and measured academic outcomes. Exploratory analyses of learning-style preferences did not yield robust effects, plausibly because text-based chatbots now operate as a near-universal default. Whether modality preferences differentiate AI engagement within a more diverse, multimodal tool ecosystem is a question for future research.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/educsci16071160/s1>. Supplementary materials present exploratory VARK modality analyses, including country-wise endorsement patterns, weak task associations, and the infeasibility of testing pre-registered modality–tool hypotheses due to ubiquitous chatbot use and minimal image-generation adoption.

Author Contributions: Conceptualization, V.D.; Methodology, V.D., E.S. and S.S.; Formal analysis, V.D. and S.S.; Investigation, V.D., E.S., A.R.S., A.A. and S.S.; Resources, A.A.; Data curation, V.D.; Writing—original draft, V.D.; Writing—review and editing, E.S., A.R.S., A.A. and S.S.; Supervision, V.D.; Project administration, E.S. All authors have read and agreed to the published version of the manuscript.

Funding: Research reported in this publication was supported by the Fogarty International Center of the National Institutes of Health under Award Number D43TW009672. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki and approved by the CUNY University Integrated Institutional Review Board on 19 September 2025 (IRB ID: 2025-0666-HCC).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available from the corresponding author upon reasonable request, subject to Institutional Review Board restrictions.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Annas, C., & Nugraha, J. (2026). Exploring the impact of generative AI literacy on students' academic performance: An ethical and academic fraud perspective. *JIBEMA: Jurnal Ilmu Bisnis, Ekonomi, Manajemen, Dan Akuntansi*, 3, 255–274. [\[CrossRef\]](#)
- Brennan, L., & Barbon, R. (2026). Flexible, VARK-aligned pre/post module assignments in a hybrid pathophysiology & pharmacology course: A mixed-methods evaluation. *Teaching and Learning in Nursing*, 21, e1145–e1149. [\[CrossRef\]](#)
- 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. [\[CrossRef\]](#)
- Chiu, T. K. (2025). AI literacy and competency: Definitions, frameworks, development and future research directions. *Journal of Educational Computing Research*, 33, 3225–3229. [\[CrossRef\]](#)
- Dong, Y., Xu, W., Huang, J., & Yann, K. (2025). Validating and refining a multi-dimensional scale for measuring AI literacy in education using the Rasch Model. *Humanities and Social Sciences Communications*, 12, 1317. [\[CrossRef\]](#)
- Francis, N. J., Jones, S., & Smith, D. P. (2025). Generative AI in higher education: Balancing innovation and integrity. *British Journal of Biomedical Science*, 81, 14048. [\[CrossRef\]](#) [\[PubMed\]](#)
- Jin, Y., Yan, L., Echeverria, V., Gašević, D., & 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. [\[CrossRef\]](#)
- Marmolejo-Ramos, F., Abadia, R., Karakale, Ö., Barrera-Causil, C., Männikkö, N., Ghorbani, B. D., Strzelecki, A., Nwizu, S., Tavares, C., & Castillo, M. (2026). University students' perceptions and adoptions of AI: A cross-national study. *Discover Artificial Intelligence*, 6, 320. [\[CrossRef\]](#)
- Marwaha, K., & Sharma, U. (2025). Debunking learning styles: Analyzing key predictors of academic success in dental education. *Advances in Physiology Education*, 49, 704–711. [\[CrossRef\]](#) [\[PubMed\]](#)
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, 100041. [\[CrossRef\]](#)

- Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles: Concepts and evidence. *Psychological Science in the Public Interest*, 9, 105–119. [\[CrossRef\]](#) [\[PubMed\]](#)
- Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching and Learning Practice*, 20, 1–24. [\[CrossRef\]](#)
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for developmental research. *Developmental Review*, 41, 71–90. [\[CrossRef\]](#) [\[PubMed\]](#)
- Ravšelj, D., Keržič, D., Tomaževič, N., Umek, L., Brezovar, N., Iahad, N. A., Abdulla, A. A., Akopyan, A., Segura, M. W., & AlHumaid, J. (2025). Higher education students' perceptions of ChatGPT: A global study of early reactions. *PLoS ONE*, 20, e0315011. [\[CrossRef\]](#) [\[PubMed\]](#)
- Sergeeva, O. V., Zheltukhina, M. R., Shoustikova, T., Tukhvatullina, L. R., Dobrokhoto, D. A., & Kondrashev, S. V. (2025). Understanding higher education students' adoption of generative AI technologies: An empirical investigation using UTAUT2. *Contemporary Educational Technology*, 17, ep571. [\[CrossRef\]](#) [\[PubMed\]](#)
- Tahir, M. S., & Zahra, A. (2026). Impact of generative artificial intelligence on student engagement, self-regulated learning, and academic integrity in higher education. *The Critical Review of Social Sciences Studies*, 4, 618–628. [\[CrossRef\]](#)
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3, 4–70. [\[CrossRef\]](#)
- World Medical Association. (2025). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human participants. *JAMA*, 333, 71–74. [\[CrossRef\]](#) [\[PubMed\]](#)
- Zou, H., Chan, K. I., Pang, P., Manditereza, B., & Shih, Y.-H. (2026). To use but not to depend: Pedagogical novelty and the cognitive brake of ethical awareness in computer science students' adoption of generative AI. *Education Sciences*, 16, 311. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.