education sciences

MDPI

Article

Changing Student Perceptions and Adoption Readiness for Generative AI Chatbots in Sino-British STEM Education: A Two-Wave Repeated Cross-Sectional Study (2023–2025)

Kamalanathan Kajan [1,*], Wenyuan Shi [2] and Dariusz Wanatowski [3]

[1] School of Electronic and Electrical Engineering, University of Leeds, Leeds LS2 9JT, UK
[2] SWJTU-Leeds Joint School, Southwest Jiaotong University, Chengdu 611756, China
[3] Faculty of Engineering and Physical Sciences, University of Leeds, Leeds LS2 9JT, UK
* Correspondence: k.kajan@leeds.ac.uk

Abstract

Generative artificial intelligence (AI) chatbots are increasingly prominent in higher education, but temporal evidence on student perceptions remains limited in English-medium transnational STEM education. This repeated cross-sectional study surveyed students at a Sino-British engineering institution in Spring 2023 (Wave 1; N = 165) and Spring 2025 (Wave 2; N = 297), and found that familiarity (37.6% vs. 67.3%; OR = 3.43), willingness (55.8% vs. 77.8%; OR = 2.78), comfort (63.6% vs. 74.4%; OR = 1.66), and perceived overall learning enhancement (46.1% vs. 69.0%; OR = 2.61) were higher in 2025. In a respondent-level model combining Wave 1 with the Spring four-programme subset (analytic N = 426), survey wave remained associated with willingness after adjustment for programme and year (OR = 3.51, 95% CI 2.17–5.69, $p < 0.001$). In the Wave 2 Spring concern checklist (N = 297), accuracy (51.5%) and deduplicated over-reliance (50.2%) were the leading concerns; because the concern blocks differed in format and options, specific concerns cannot be shown to have increased or matured. The findings indicate higher population-level adoption readiness, not individual progression or routine use. "Concern maturation" remains a hypothesis requiring invariant instruments and panel designs.

Keywords: AI chatbots; generative AI; adoption readiness; student perceptions; transnational education; STEM; repeated cross-sectional; Diffusion of Innovations

1. Introduction

Generative AI (GenAI) chatbots have become widely visible across higher education, with large surveys and multi-country studies reporting substantial student engagement alongside persistent concerns about trustworthiness, over-reliance, and academic integrity (Freeman, 2025; Attewell, 2025; Ravšelj et al., 2025; Cabrera & Neville, 2025). Nevertheless, much of the evidence remains cross-sectional. Temporal studies are still uncommon, and familiarity or willingness should not be assumed to represent actual frequency or purposes of use; for example, Polyportis (2024) reports that usage can decline even while a technology becomes more widely known. This distinction is important because the present surveys measure perceptions and adoption readiness rather than behavioural normalisation.

A second gap concerns context. English-medium instruction (EMI) transnational education (TNE) programmes in China combine UK-aligned curricula with local learning cultures and a mainland digital environment in which access to commercial platforms may differ

Check for updates

Academic Editors: Alexandra Costa and Marina Sousa

Received: 21 July 2026
Revised: 19 August 2026
Accepted: 21 August 2026
Published: 24 August 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) license.

from UK settings. These programmes are underrepresented in research on GenAI adoption, limiting evidence for context-sensitive guidance, assessment, and student support.

This paper addresses these gaps using samples drawn from the same Sino-British joint engineering institution. The published single-institution Wave 2 study reported item-level descriptive findings from 297 respondents, including adoption readiness, perceived benefits, concerns, and requested features (Kajan et al., 2025); a later multi-institution study used the broader 2025 dataset (N = 467, including 300 respondents from the present institution) to examine lecturer-role expectations, social ambivalence, and ambivalence literacy (Kajan et al., 2026b); and a related conference paper examined integration preferences using the broader 2025 dataset (Kajan et al., 2026a). The present paper introduces the previously unreported Wave 1 baseline (Spring 2023; N = 165) and contributes new between-wave comparisons of common items and a respondent-level adjusted analysis. Neither the 2023 baseline nor the 2023–2025 temporal comparisons were reported in the earlier publications. Appendix D summarises the relationship among these outputs.

The paper gives evidential priority to directly comparable measures of familiarity, importance, willingness, comfort, and perceived learning enhancement. Concern and desired-support items are reported as separate wave-specific profiles because Wave 1 used a restricted single-choice format whereas Wave 2 used broader multiple-response checklists. Diffusion of Innovations (DOI; Rogers, 2003) is used as a sensitising framework rather than as a validated stage model. "Concern maturation" is considered a possible interpretation generated by the Wave 2 profile, not an empirically confirmed temporal process.

Research Questions

RQ1. How did student familiarity, perceived importance, willingness, comfort, and perceived learning enhancement differ between the 2023 and 2025 samples?

RQ2. Did the association between survey wave and willingness remain after adjustment for programme and year of study?

RQ3. What concerns and desired forms of support were reported in each wave, taking account of differences in item content and response format?

RQ4. How might the observed patterns be cautiously interpreted through Diffusion of Innovations theory in a Sino-British transnational STEM context?

Programme- and year-level analyses are treated as exploratory because several subgroup cells are small and the Wave 2 Year 4 cell contains only two students.

2. Literature Review

2.1. Temporal Evidence on Student AI Adoption

Large-scale studies document rapid growth in awareness and reported use of GenAI in higher education (Freeman, 2025; Attewell, 2025; Ravšelj et al., 2025), and reviews also identify methodological fragmentation, particularly a predominance of single-timepoint surveys (Schei et al., 2024; Albadarin et al., 2024; Abdallah et al., 2025). Emerging temporal studies address different questions; for instance, Polyportis (2024) followed the same students and examined behavioural change; Parker et al. (2026) used repeated cross-sectional samples to describe population-level differences; and Ma et al. (2026) examined successive cohorts in an introductory programming context. These studies show why panel and repeated cross-sectional designs should not be treated as interchangeable: panels estimate within-person trajectories, whereas repeated cross-sections estimate differences between sampled populations at different historical moments.

The present study extends this temporal literature to an EMI Sino-British STEM setting. Its originality lies less in the use of DOI or repeated cross-sectional design than in

the institutional context and in combining directly comparable readiness measures with clearly bounded, wave-specific profiles of concerns and support preferences.

2.2. Perceived Educational Value and Student Concerns

Research in STEM suggests that AI chatbots may support explanation, feedback, conceptual clarification, and problem-solving, although effects depend on educational level, intervention design, discipline, and duration (Deng et al., 2025; Z. Liu et al., 2025; Heung & Chiu, 2025; El Fathi et al., 2025; Bravo & Cruz-Bohorquez, 2024). The present study does not measure objective attainment: its enhancement items capture perceived educational value only, and stability in collaboration or creativity ratings cannot establish that students' expectations are objectively calibrated.

Student concerns commonly include reduced human interaction, generic feedback, accuracy, dependency, academic integrity, and privacy or governance issues (Thomson et al., 2024; Farhat, 2025; Park & Ahn, 2024; Ortiz-Bonnin & Blahopoulou, 2025; Wu & Carroll, 2026). It is plausible that the salience of these concerns changes as the technological and institutional environment changes. However, such development can only be tested directly when the same constructs are measured equivalently across time. The present data therefore support description of each wave's concern profile, while any developmental interpretation remains provisional.

2.3. Transnational and English-Medium Context

Students in EMI-TNE programmes work across discipline-specific language, UK-aligned academic conventions, and a mainland Chinese technology environment; consequently, language and terminology support may be especially relevant. The current study did not measure English proficiency and has no non-TNE comparison group; it therefore cannot establish that any language-support pattern is distinctive to TNE, but it can identify a pattern that warrants targeted comparative research. Related evidence from Chinese higher education suggests that critical thinking and digital efficacy can shape AI acceptance and willingness (C. Liu et al., 2025; Wang & Li, 2024), although these mechanisms were not measured in the present study.

3. Theoretical Framework

Rogers' (2003) DOI framework describes diffusion through knowledge, persuasion, decision, implementation, and confirmation, shaped by perceived relative advantage, compatibility, complexity, trialability, and observability. DOI is useful for organising questions about how perceptions may differ as a technology becomes more established. However, the present survey was not designed as a validated operationalisation of Rogers' stages or attributes; familiarity, willingness, comfort, perceived enhancement, concerns, and desired features have conceptual relevance to diffusion, but they do not directly measure knowledge, compatibility, observability, or confirmation.

Accordingly, DOI is used here as an interpretive or sensitising framework. The repeated cross-sectional design estimates population-level differences between two samples; it does not show that individual students, a single cohort, or the institution progressed through specified DOI stages. A panel design would answer a different and complementary question by examining within-person change.

DOI describes adoption processes, but does not itself predict the content of user concerns; therefore, the proposition below is our extension of diffusion thinking rather than a prediction contained in Rogers' framework. We retain the phrase "concern maturation" only as a proposition: as familiarity with GenAI increases, concerns may become more focused on practical consequences such as dependency, accuracy, and integrity. The present

study cannot test that developmental mechanism because the concern instruments were not invariant. Evidence consistent with the proposition should therefore be treated as hypothesis-generating and requires validation using identical instruments, validated scales, behavioural usage measures, longitudinal panels, and cross-institutional comparison.

4. Methods

4.1. Research Design and Setting

We conducted a two-wave repeated cross-sectional survey at an English-medium Sino-British joint engineering institution in China: Wave 1 was administered online between 31 March and 10 May 2023, and the SWJTU-Leeds Wave 2 component was administered between 2 and 28 April 2025. The broader four-institution survey from which it was drawn was conducted between March and December 2025 (Kajan et al., 2026b); however, only SWJTU-Leeds Joint School respondents are considered in this two-wave study. Teaching and assessment at the institution follow UK-aligned standards within a dual-degree framework.

4.2. Participants and Analysis Samples

As part of this revision, the published Wave 2 aggregates were checked against the recovered Spring-cohort records. Some values reproduce exactly, including perceived importance (259/297), overall learning enhancement (205/297), and the two language-related item counts. The recovered Spring cohort contained 50 voice-based endorsements (50/297, 16.8%), matching the count presented graphically in the earlier article (Kajan et al., 2025), but not its narrative percentage. Six published aggregates could not be reproduced from the recovered records or from the original survey-platform export under the stated top-two coding rule: the recovered records give 200/297 (67.3%) for familiarity, 231/297 (77.8%) for willingness, 221/297 (74.4%) for comfort, 142/297 (47.8%) for problem-solving, 106/297 (35.7%) for collaboration, and 103/297 (34.7%) for creativity. For the present analysis, the recovered Spring 2025 respondent-level records are treated as the authoritative source for the directly comparable Wave 2 outcomes because they permit respondent-level verification and reproducible calculation of numerators, denominators, and effect estimates; a provenance audit of the original survey-platform exports could not reconstruct the tabulation that produced the earlier aggregates. The previously published values (199/297 familiarity, 232/297 willingness, 235/297 comfort, 143/297 problem-solving, 104/297 collaboration, and 98/297 creativity) are reported only to document the provenance discrepancy. Recalculation using the recovered records changes several descriptive values and effect estimates but does not alter the substantive inferential conclusions.

The original Wave 1 respondent-level dataset was recovered. The primary adjusted analysis was fitted directly at respondent level using all 165 Wave 1 records and 261 Spring 2025 records from the four programmes represented in both waves: Civil Engineering with Transport (CET), Electronic and Electrical Engineering (EEE), Computer Science (CS), and Mechanical Engineering (ME). The 36 Wave 2 Materials Science and Engineering records were excluded because that programme was not represented in Wave 1. No records in the primary model were missing programme, year-of-study, or willingness data, giving an analytic sample of $N = 426$. A robustness re-estimation combined the 165 Wave 1 records with the archived four-programme subset of 262 records, giving $N = 427$. That archived subset includes two later CET respondents and excludes one Spring CS record; its wave estimate was materially identical to that of the primary model. Table 1 summarises the participant characteristics of the primary descriptive cohorts, while Table 2 sets out the derivation, analytical use, and limitations of each analysis sample.

Table 1. Participant characteristics in the primary descriptive cohorts.

Characteristic	Wave 1 (2023), $N = 165$	Wave 2 (2025), $N = 297$	Note
Gender	Not collected	Male 238 (80.1%); Female 59 (19.9%)	No between-wave gender comparison
CET	68 (41.2%)	40 (13.5%)	Represented in both waves
EEE	35 (21.2%)	119 (40.1%)	Represented in both waves
CS	32 (19.4%)	66 (22.2%)	Represented in both waves
ME	30 (18.2%)	36 (12.1%)	Represented in both waves
MSE	Not present	36 (12.1%)	Wave 2 only
Year 1	75 (45.5%)	158 (53.2%)	
Year 2	33 (20.0%)	48 (16.2%)	
Year 3	38 (23.0%)	89 (30.0%)	
Year 4	19 (11.5%)	2 (0.7%)	Wave 2 percentage not used for inference

Note. Wave 2 programme and year counts are tabulated from the recovered Spring respondent-level records. The published article reported 38 for Civil Engineering with Transport, 65 for Computer Science, 39 for Mechanical Engineering, 157 for Year 1, and 3 for Year 4; the differences are documented in Section 4.2. Gender was not recorded in the recovered export, so the published gender distribution is reproduced and could not be re-verified.

Table 2. Data provenance and uses of each analysis sample.

Dataset	Derivation	*N*	Use in This Paper	Limitation
Wave 1	Original respondent-level records; complete programme, year, and willingness data	165	Primary descriptive comparisons and respondent-level adjusted model	Invitation denominator not retained
Recovered Spring Wave 2	Consenting Spring 2025 responses (completed 2–28 April 2025)	297	Primary Wave 2 source for programme/year characteristics, directly comparable outcomes, concern profiles, desired-support profiles, and the Spring four-programme subset	Gender was unavailable in the recovered export; six outcome values differ from the earlier published tabulation (Section 4.2)
Archived Wave 2 export	296 Spring records carrying assigned SWJTU-Leeds labels in the archived multi-institution export, plus four later SWJTU-Leeds respondents collected in November–December 2025; one Spring CS record retained a blank institution field	300	Provenance documentation of the multi-institution archive	Construction not directly reproducible from raw export fields (Section 4.2)
Archived Wave 2 overlap subset	CET, EEE, CS, and ME records in the archived export; excludes Wave 2-only MSE ($n = 36$) and Environmental Engineering ($n = 2$)	262	Robustness re-estimation of the adjusted model ($N = 427$)	Includes two later CET respondents; does not capture one Spring CS record with a blank institution field
Primary combined adjusted model	Wave 1 (165) + Spring four-programme subset (261)	426	Primary respondent-level willingness regression	Excludes the Wave 2-only MSE programme; archived-subset robustness re-estimation reported in Section 5.2

4.3. Instrument and Between-Wave Comparability

Both waves included core items measuring familiarity, perceived importance, willingness, comfort, and perceived enhancement of overall learning, problem-solving, collaboration, and creativity. Familiarity, willingness, comfort, and the four perceived-enhancement items used identical five-point response structures. The importance item also used identical response categories in both waves—Not important, Neutral, Important, and Very important—and is therefore treated as directly comparable rather than as a harmonised non-equivalent item. Comparability was verified by inspecting the unique response labels and full frequency distributions in the original Wave 1 and recovered Wave 2 respondent-level exports; both files contained exactly these four categories, with no additional category or retrospective harmonisation (Appendix A.1).

The concern and desired-feature blocks were not equivalent. Wave 1 largely required a single response from a short menu, with an additional field, whereas Wave 2 allowed multiple selections from a fixed concern checklist, whose recovered respondent-level export contains 15 distinct label strings representing 14 conceptual concerns (the over-reliance wording appears in two variants differing only by a terminal full stop; the two labels were combined at the respondent level for prevalence reporting) and 14 fixed feature-response options, comprising 13 substantive requested features and an "I have no specific preferences" option; an optional "other" text field was also available for each block. In Wave 1, response options competed with each other, while in Wave 2, several options could be endorsed; consequently, checklist percentages do not estimate the same response process and are not used to claim increases, decreases, emergence, replacement, or addition across waves. Appendix A reproduces the wording and options preserved in the recovered respondent-level exports. Table 3 summarises the degree of comparability between the two waves and the interpretations permitted for each survey block.

Table 3. Comparability and permitted interpretation of survey blocks.

Construct	Wave 1	Wave 2	Comparability	Permitted Interpretation
Familiarity	Five-point Likert	Five-point Likert	Direct	Between-wave comparison
Willingness	Five-point Likert	Five-point Likert	Direct	Between-wave comparison
Comfort	Five-point Likert	Five-point Likert	Direct	Between-wave comparison
Perceived enhancement	Five-point Likert	Five-point Likert	Direct	Between-wave comparison of perceptions
Importance	Identical four-category ordinal scale	Identical four-category ordinal scale	Direct	Between-wave comparison
Concerns	Restricted single choice + Other	15 exported select-all label strings (14 concerns; over-reliance in two variants) + optional Other response	Not equivalent	Separate wave-specific profiles
Desired features	Restricted single choice + Other	14 fixed select-all options: 13 substantive features plus "I have no specific preferences"; optional Other response	Not equivalent	Separate wave-specific profiles

4.4. Data Collection, Ethics, and De-Identification

Participation in both waves was voluntary and unrelated to assessment, and survey invitations were distributed through course channels. Invitation records were not retained, so exact response rates cannot be calculated. Self-selection may have favoured students with greater interest in or experience of AI.

Wave 1 was conducted as a pilot study reviewed under the ethical review processes of the Faculty of Engineering and Physical Sciences at the University of Leeds (University of Leeds Code of Research Conduct and Research Ethics; no individual reference issued for pilot reviews). Wave 2 and the two-wave comparative design as a whole were approved by the Faculty of Engineering and Physical Sciences Research Ethics Committee, University of Leeds (Ref. 2714; 8 March 2025). Both waves were conducted in accordance with the Declaration of Helsinki. Participants in Wave 1 were informed that participation was voluntary, that they could withdraw and request that their data not be used, and that responses would be treated as confidential and reported in aggregate. The raw Wave 1 administrative export contained names and email addresses; these identifiers were removed before the present analysis, and no linkage key was retained. No personally identifying information was collected in Wave 2 (Kajan et al., 2025).

The present study uses de-identified data and aggregate reporting. The slight compositional difference between the archived Wave 2 export and the recovered Spring cohort is reported as a data-governance limitation (Section 7.2).

4.5. Statistical Analysis

Frequencies and Wilson 95% confidence intervals were calculated for proportions. For the directly comparable items, responses were dichotomised into their upper two categories to reproduce the planned and previously reported comparison thresholds and to provide interpretable odds ratios. For willingness, "Somewhat likely" and "Very likely" were coded as 1, while "Neutral", "Somewhat unlikely", and "Very unlikely" were coded as 0. This dichotomisation reduces information from the original ordinal scales and is acknowledged as a limitation.

Between-wave differences were evaluated using Pearson chi-square tests and odds ratios. Benjamini–Hochberg false discovery rate correction was applied separately to (a) the four familiarity, importance, willingness, and comfort comparisons; (b) the four perceived-enhancement comparisons; (c) the eight exploratory programme and year comparisons. The sparse Wave 2 Year 4 result was included in the multiplicity family but was not interpreted.

A respondent-level logistic regression predicted top-two willingness from survey wave, programme, and year of study. The primary model included all 165 Wave 1 respondents and 261 Spring Wave 2 respondents from CET, EEE, CS, and ME, giving an analytic sample of $N = 426$. Wave 2023, EEE, and Year 1 were the reference categories. Because respondent-level records were available for both waves, the model was fitted directly to individual observations rather than to grouped frequency counts. A wave × programme interaction model was fitted as an exploratory analysis of possible programme-level heterogeneity. Model assessment used the Akaike Information Criterion, Nagelkerke R^2, likelihood-ratio comparison, confidence intervals, and coefficient stability. An archived-subset model using $N = 427$ was fitted as a robustness re-estimation. The Hosmer–Lemeshow test was not retained as the principal model diagnostic because the categorical predictors produced tied fitted probabilities and its result was sensitive to the grouping procedure.

The respondent-level logistic regression and verification analyses were conducted in Python 3.13.5 using NumPy 2.3.5, SciPy 1.17.0, and statsmodels 0.14.6. All item-level

comparisons and Benjamini–Hochberg corrections for the directly comparable outcomes were recomputed from the recovered respondent-level records.

5. Results

5.1. Directly Comparable Adoption-Readiness Outcomes (RQ1)

Table 4 presents the between-wave comparisons for familiarity, perceived importance, willingness, and comfort.

Table 4. Adoption readiness and importance comparison.

Item	Wave 1 *n*/*N* (%)	Wave 2 *n*/*N* (%)	Difference (pp)	OR (95% CI)	*p*; BH q
Familiar/very familiar	62/165 (37.6%)	200/297 (67.3%)	+29.8	3.43 (2.30–5.10)	$p < 0.001$; q < 0.001
Important/very important	142/165 (86.1%)	259/297 (87.2%)	+1.1	1.10 (0.63–1.93)	$p = 0.728$; q = 0.728
Somewhat/very willing	92/165 (55.8%)	231/297 (77.8%)	+22.0	2.78 (1.84–4.19)	$p < 0.001$; q < 0.001
Somewhat/very comfortable	105/165 (63.6%)	221/297 (74.4%)	+10.8	1.66 (1.10–2.51)	$p = 0.015$; q = 0.020

Familiarity and willingness were substantially higher, and comfort was moderately higher, in the 2025 sample, and perceived importance was high and statistically similar in both samples. Although importance and willingness showed different descriptive patterns, these results should not be interpreted as confirmation that students or the institution progressed between specified DOI stages. Table 5 presents the corresponding between-wave comparisons for perceived learning enhancement.

Table 5. Perceived learning enhancement (top two categories: Very much/Extremely).

Domain	Wave 1 *n*/*N* (%)	Wave 2 *n*/*N* (%)	Difference (pp)	OR (95% CI)	*p*; BH q
Overall learning	76/165 (46.1%)	205/297 (69.0%)	+23.0	2.61 (1.76–3.86)	$p < 0.001$; q < 0.001
Problem-solving	63/165 (38.2%)	142/297 (47.8%)	+9.6	1.48 (1.01–2.19)	$p = 0.046$; q = 0.092
Collaboration	58/165 (35.2%)	106/297 (35.7%)	+0.5	1.02 (0.69–1.52)	$p = 0.908$; q = 0.908
Creativity	61/165 (37.0%)	103/297 (34.7%)	−2.3	0.91 (0.61–1.35)	$p = 0.622$; q = 0.830

Perceived overall learning enhancement was higher in Wave 2; the problem-solving comparison did not remain significant after correction, and collaboration and creativity ratings were similar across waves, indicating a differentiated pattern of perceived utility rather than demonstrated educational effects or objectively calibrated expectations.

For transparency, recomputing the same comparisons using the previously published aggregate values leaves every inferential conclusion unchanged after false-discovery-rate correction (for comfort, OR = 2.17, 95% CI 1.42–3.31, q < 0.001), and the problem-solving difference is non-significant after correction under either set of counts (published q = 0.078; recovered q = 0.092). Figure 1 provides a visual summary of the adoption-readiness and perceived-enhancement comparisons across the two waves, with Wilson 95% confidence intervals.

5.2. Adjusted Association with Willingness (RQ2)

In the primary respondent-level model, using 165 Wave 1 and 261 Spring Wave 2 respondents from the four shared programmes (analytic $N = 426$), Wave 2 was associated with higher odds of willingness after adjustment for programme and year of study (OR = 3.51, 95% CI 2.17–5.69, $p < 0.001$). The main-effects model had an AIC of 499.92 and a Nagelkerke R^2 of 0.122. An exploratory wave × programme model improved AIC to 497.09, and the likelihood-ratio comparison was significant ($\chi^2(3) = 8.83$, $p = 0.032$), indicating that the magnitude of the wave association may differ by programme. The

corresponding programme-specific Wave 2 vs. Wave 1 odds ratios were 1.25, 4.07, 3.74, and 10.33 for CET, EEE, CS, and ME, respectively. These interaction findings are exploratory because several subgroup cells are small and the associated uncertainty is wide. The model reduces measured programme and year compositional differences but cannot separate historical change from unmeasured cohort, institutional, or technological differences. The model uses the recovered Spring records directly, so its cohort matches the descriptive $N = 297$ comparisons.

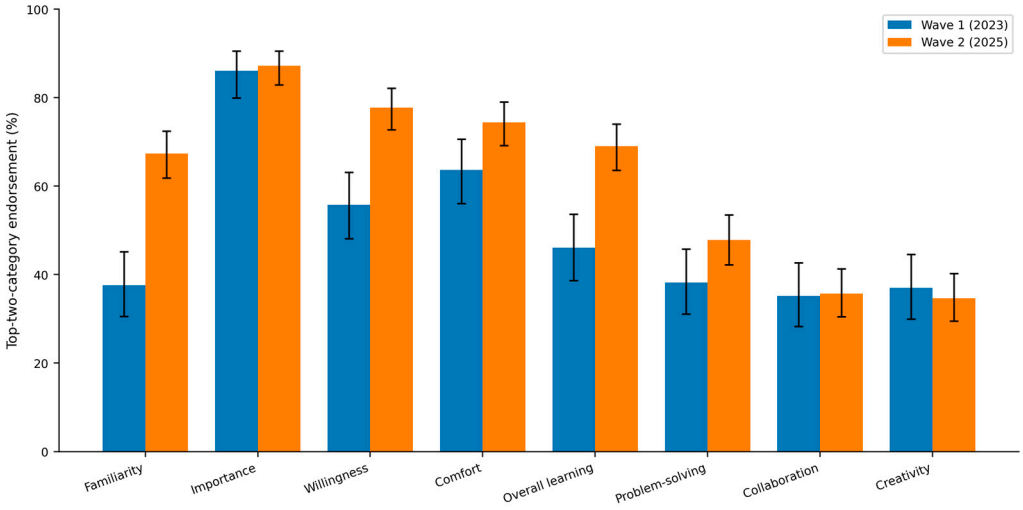


Figure 1. Adoption-readiness and perceived-enhancement indicators in the two samples. Error bars show Wilson 95% confidence intervals. The importance item used identical four-category response scales in both waves.

A robustness re-estimation using the archived-export four-programme subset ($N = 427$) produced a materially identical wave estimate (OR = 3.55, 95% CI 2.19–5.73), and a separate unadjusted sensitivity analysis restricted to EEE, CS, and ME retained a large association between survey wave and willingness (Wave 1 $n = 97$; Spring Wave 2 $n = 221$; OR = 4.71, 95% CI 2.81–7.90). These checks support the stability of the adjusted association but cannot rule out unmeasured differences between the samples.

5.3. Wave-Specific Concern Profiles (RQ3)

Table 6 presents the two concern blocks separately: Wave 1 percentages are shares of a single-choice distribution, and Wave 2 percentages are multiple-response endorsement rates in the recovered Spring cohort ($N = 297$). They are not estimates of the same response process and are not compared inferentially.

Table 6. Concern profiles by wave; response formats are not directly comparable.

Wave/Format	Concern	*n*/*N* (%)	Interpretive Status
Wave 1: single choice	Lack of human interaction	54/165 (32.7%)	Restricted menu
Wave 1: single choice	Lack of personalised feedback	53/165 (32.1%)	Restricted menu
Wave 1: single choice	Inability to ask follow-up questions	29/165 (17.6%)	Restricted menu
Wave 1: single choice	Difficulty understanding responses	16/165 (9.7%)	Restricted menu
Wave 1: other text	Other free-text concerns (all)	13/165 (7.9%)	Completes the single-choice distribution

Table 6. *Cont.*

Wave/Format	Concern	*n*/*N* (%)	Interpretive Status
Wave 1: other text	Selected examples within "Other": accuracy/error (4); over-reliance/dependency (1)	Subset of 13/165	Illustrative coding; not additional responses
Wave 2: select all	Incorrect or unclear answers	153/297 (51.5%)	Option introduced in Wave 2 checklist
Wave 2: select all	Over-reliance/learning dependency	149/297 (50.2%)	Deduplicated across two exported label variants (see table note)
Wave 2: select all	Lack of human interaction	125/297 (42.1%)	Shared theme; different response process
Wave 2: select all	May not understand questions clearly	118/297 (39.7%)	Wave 2 option
Wave 2: select all	Lack of personalised feedback	96/297 (32.3%)	Shared theme; different response process
Wave 2: select all	Academic integrity/plagiarism	89/297 (30.0%)	Wave 2 option
Wave 2: select all	Unable to ask follow-up questions	76/297 (25.6%)	Shared theme; different response process
Wave 2: select all	Struggles with advanced topics	76/297 (25.6%)	Wave 2 option
Wave 2: select all	Answers lack depth/detail	74/297 (24.9%)	Wave 2 option
Wave 2: select all	Difficulty understanding responses	53/297 (17.8%)	Shared theme; different response process
Wave 2: select all	Reduced motivation/interest	44/297 (14.8%)	Wave 2 option
Wave 2: select all	Technical/internet problems	36/297 (12.1%)	Wave 2 option
Wave 2: select all	Translation misunderstanding	35/297 (11.8%)	Wave 2 option
Wave 2: select all	Language barriers in English	15/297 (5.1%)	Wave 2 option

Note. The recovered Wave 2 respondent-level export contained two over-reliance label variants that differed only by a terminal full stop. In the Spring cohort, 134 respondents selected the variant with the terminal full stop, 66 selected the variant without it, and 51 selected both. After combining the variants at the respondent level, the prevalence in the Spring cohort was 149/297 (50.2%). The undeduplicated total of 200 endorsements counted the 51 overlapping respondents twice.

Six Wave 2 respondents also used the optional "Other" concern field; these heterogeneous free-text entries were retained for provenance but were not included in the fixed-option frequencies.

The earlier article reported over-reliance at 67%, content quality/accuracy at 52%, and reduced human interaction at 42% (Kajan et al., 2025). After respondent-level deduplication in the Spring cohort, incorrect or unclear answers (153/297; 51.5%) and over-reliance/learning dependency (149/297; 50.2%) were the two most frequently endorsed concerns, followed by reduced human interaction (125/297; 42.1%). Table 6 reports these exact, traceable Spring-cohort counts; the earlier 67% value is cited only to explain the discrepancy arising from the two exported label variants. The Wave 2 profile shows high endorsement of practical and epistemic concerns, but does not establish that these concerns increased after, replaced, or were added to Wave 1 concerns, because the relevant options were not measured equivalently.

5.4. Wave-Specific Desired-Support Profiles (RQ3)

Table 7 presents the desired-support profiles separately for each wave, reflecting the different response formats.

Table 7. Desired-support profiles by wave; response formats are not directly comparable.

Wave/Format	Desired Feature or Support	*n*/*N* (%)
Wave 1: single choice	Personalised responses	57/165 (34.5%)
Wave 1: single choice	Ability to ask follow-up questions	56/165 (33.9%)
Wave 1: single choice	Choose chatbot personality	31/165 (18.8%)
Wave 1: single choice	Use for assignments/assessments	20/165 (12.1%)
Wave 1: other text	Other free-text feature	1/165 (0.6%)
Wave 2: select all	Help with technical words/language	154/297 (51.9%)
Wave 2: select all	Ability to ask follow-up questions	145/297 (48.8%)
Wave 2: select all	Personalise responses	129/297 (43.4%)
Wave 2: select all	Step-by-step problem-solving guides	107/297 (36.0%)
Wave 2: select all	Translate/simplify explanations	101/297 (34.0%)
Wave 2: select all	Links to extra study resources	97/297 (32.7%)
Wave 2: select all	Interactive visuals/animations	80/297 (26.9%)
Wave 2: select all	Choose chatbot personality	75/297 (25.3%)
Wave 2: select all	Assignment/assessment preparation	70/297 (23.6%)
Wave 2: select all	Explain complex English vocabulary	68/297 (22.9%)
Wave 2: select all	Chinese–English responses	66/297 (22.2%)
Wave 2: select all	Voice-based questions	50/297 (16.8%)
Wave 2: select all	Group/team interaction support	49/297 (16.5%)
Wave 2: select all	I have no specific preferences	13/297 (4.4%)

Note. Wave 2 counts in this table are calculated from the recovered Spring 2025 cohort (N = 297), which is the authoritative Wave 2 source for the present paper. Two respondents used the optional "Other" feature field; both entries were non-substantive or uncodable and were not included in the fixed-option frequencies.

The published Wave 2 paper described voice-based Q&A narratively at approximately 23% (Kajan et al., 2025). The recovered Spring cohort contained 50 endorsements (50/297, 16.8%), matching the count presented graphically in the earlier article (Kajan et al., 2025), but not its narrative percentage.For language support, the published N = 297 paper reported that 52.0% requested language/terminology support, 5.1% selected the English-language-barrier item, and 11.8% selected translation misunderstanding (Kajan et al., 2025). The paper also reported a combined item-level language-related figure of 16.8%; because the question permitted multiple selections, this figure should not be interpreted as the proportion of unique students selecting at least one of the two items. In the recovered Spring cohort, 15 students selected the English-language-barrier item, 35 selected translation misunderstanding, and five selected both; therefore, 45/297 students (15.2%) selected at least one of the two language-related concerns.

Language/terminology support was selected by 154/297 students (51.9%), producing descriptive differences of 46.8 percentage points relative to the narrow language-barrier item and 36.7 percentage points relative to the unique-student broader measure. Using the rounded figures reported in the published paper gives differences of 46.9 and 35.2 percentage points, respectively, although the latter uses its combined item-level figure rather than a deduplicated prevalence estimate. These contrasts involve different constructs and should not be interpreted as precise estimates of unmet need. Without English-proficiency measures or a non-TNE comparison group, the pattern cannot be classified as distinctive to TNE.

5.5. Exploratory Programme and Year Patterns

Table 8 presents the exploratory willingness comparisons by programme and year of study within the four-programme respondent-level subset.

Table 8. Exploratory willingness comparisons in the four-programme row-level subset.

Subgroup	Wave 1 *n*/*N* (%)	Wave 2 *n*/*N* (%)	Difference (pp)	OR (95% CI)	*p*; BH q
CET	48/68 (70.6%)	29/40 (72.5%)	+1.9	1.10 (0.46–2.62)	$p = 0.832$; q = 0.832
EEE	16/35 (45.7%)	92/119 (77.3%)	+31.6	4.05 (1.83–8.93)	$p < 0.001$; q = 0.001
CS	19/32 (59.4%)	55/66 (83.3%)	+24.0	3.42 (1.31–8.91)	$p = 0.010$; q = 0.019
ME	9/30 (30.0%)	29/36 (80.6%)	+50.6	9.67 (3.10–30.11)	$p < 0.001$; q < 0.001
Year 1	28/75 (37.3%)	104/127 (81.9%)	+44.6	7.59 (3.96–14.54)	$p < 0.001$; q < 0.001
Year 2	24/33 (72.7%)	37/43 (86.0%)	+13.3	2.31 (0.73–7.33)	$p = 0.148$; q = 0.237
Year 3	28/38 (73.7%)	63/89 (70.8%)	−2.9	0.87 (0.37–2.03)	$p = 0.740$; q = 0.832
Year 4	12/19 (63.2%)	1/2; percentage suppressed	Not interpreted	Not interpreted	$p = 0.716$; q = 0.832; not interpreted

Note. The Wave 2 Year 4 group contained only two respondents; therefore, the percentage, odds ratio, and substantive comparison are not interpreted. The test was retained conservatively within the defined eight-test multiplicity family.

Programme heterogeneity in willingness was statistically evident in Wave 1 but not in the Spring Wave 2 four-programme subset (Wave 1: Cramér's V = 0.31, $\chi^2(3) = 15.73$, $p = 0.001$; Wave 2: V = 0.09, $\chi^2(3) = 1.96$, $p = 0.581$). Year heterogeneity followed a similar descriptive pattern (Wave 1: V = 0.34, $\chi^2(3) = 19.54$, $p < 0.001$; Wave 2: V = 0.16, $\chi^2(3) = 6.42$, $p = 0.093$). These subgroup analyses are exploratory and do not justify assigning programmes to DOI stages or attributing differences to EMI adjustment, curriculum, or accumulated experience.

6. Discussion

6.1. Population-Level Differences in Adoption Readiness

The strongest evidence concerns identical measures. The 2025 sample reported higher familiarity, willingness, comfort, and perceived overall learning enhancement. In the primary respondent-level adjusted model, the wave association with willingness remained after adjustment for programme and year. The model reduces measured programme and year compositional differences, but survey wave still combines historical, cohort, and institutional differences and changes in the wider technological environment that the design cannot separate.

One component of that changed environment deserves explicit attention. In Spring 2023, direct access to the ChatGPT service from mainland China was restricted, and users often relied on indirect access arrangements (Reuters, 2023). ChatGPT is mentioned here only to describe the platform-access environment; no specific ChatGPT model was used as a study intervention, and the model versions accessed by respondents were not recorded. By Spring 2025, domestically developed generative AI systems had become much more widely available; a subsequent CNNIC report estimated approximately 249 million generative AI users in China by December 2024, rising to 515 million by June 2025, with domestically developed models broadly popular among users (Xinhua, 2025). Part of the between-wave increase in familiarity, willingness, and comfort may therefore reflect the removal of practical access barriers rather than attitudinal or stage-like progression in the diffusion-theoretic sense. The design cannot apportion the observed differences

between access- and perception-driven change, and this qualification applies directly to the DOI-informed interpretation.

These results are best described as higher population-level adoption readiness in 2025; they do not demonstrate that the same students progressed, that the institution moved through a DOI stage, or that chatbot use became routine. The surveys did not measure actual frequency, duration, or purposes of use. This distinction is reinforced by Polyportis (2024), whose behavioural panel findings show that familiarity and willingness need not move in parallel with usage.

Perceived overall learning enhancement was higher in the 2025 sample, whereas collaboration and creativity ratings were broadly similar across samples, suggesting differentiated perceived utility, especially for individual learning support. There is no evidence that AI improved objective learning outcomes or that students' expectations were accurate.

6.2. Concern Profiles and the Concern-Maturation Proposition (RQ3–RQ4)

Accuracy, deduplicated over-reliance, and reduced human interaction were the most frequently endorsed concerns in the Wave 2 checklist; academic integrity was also selected by 30.0% of respondents. One hypothesis is that the high endorsement of practical and epistemic concerns in the higher-familiarity 2025 sample reflects greater attention to consequences of AI use. However, substantive historical change and measurement artefact are empirically inseparable here. Wave 1 required a single choice from a limited menu, Wave 2 permitted several endorsements from a much broader list, and the two leading Wave 2 items were not available as equivalent Wave 1 options.

DOI can support the plausibility of a developmental proposition, but it cannot compensate for absent measurement equivalence. We therefore do not claim that concerns matured, emerged, increased, replaced earlier concerns, or were additive; "concern maturation" remains a hypothesis for future studies with invariant instruments and within-person or repeated population measures.

6.3. Language and Terminology Support in EMI-TNE

In the recovered Spring cohort, 154/297 students (51.9%) requested language/terminology support, while 15/297 (5.1%) selected English-language barriers and 35/297 (11.8%) selected translation misunderstanding. After accounting for the five respondents who selected both concern items, 45/297 students (15.2%) selected at least one language-related concern. This pattern may be especially relevant in an EMI-TNE setting. However, feature demand and concern endorsement are distinct constructs, and the absence of English-proficiency measures and a non-TNE comparison group means that this pattern cannot be treated as distinctive to TNE; instead, it identifies a priority for comparative research.

6.4. Institutional Implications

The findings suggest priorities for institutional consideration, not interventions whose effectiveness has been tested. These priorities include explicit verification and source-checking practices, transparent disclosure of AI use, process-oriented assessment, guidance on cognitive dependency, and optional language/terminology scaffolds. AI-literacy and assessment frameworks may help institutions design and evaluate such approaches (Chiu et al., 2024; Chee et al., 2025; Khlaif et al., 2025; Bearman et al., 2024; Kofinas et al., 2025). Future evaluations should measure behavioural use, learning outcomes, integrity, and equity rather than infer effectiveness from perception data.

7. Conclusions, Limitations, and Future Research

7.1. Conclusions

The 2025 sample reported higher familiarity, willingness, comfort, and perceived overall learning enhancement regarding GenAI chatbots than the 2023 sample. In the primary respondent-level adjusted model, wave remained associated with willingness after accounting for programme and year of study. These results indicate higher population-level adoption readiness in the 2025 sample but do not demonstrate individual progression, behavioural normalisation, or causal effects of accumulated experience.

Accuracy, deduplicated over-reliance/learning dependency, and reduced human interaction were prominent in the Wave 2 concern profile, and academic integrity was also frequently endorsed. Because the concern blocks differed in content and response process, the study cannot establish that these concerns increased, matured, supplemented, or replaced earlier concerns, and concern maturation is retained only as a hypothesis. Institutional attention to responsible reliance, verification, integrity, assessment processes, and context-sensitive language support is reasonable, but these approaches require direct evaluation.

7.2. Limitations

i. The repeated cross-sectional design compares different samples and cannot identify within-person change.
ii. The survey wave combines historical, cohort, institutional, and technological differences that cannot be disentangled, including changes in the practical accessibility of generative AI platforms in mainland China between 2023 and 2025.
iii. The surveys measure perceptions and adoption readiness rather than actual usage frequency, duration, or purposes.
iv. The concern and desired-feature blocks were not equivalent: Wave 1 used a restricted single-choice format, whereas Wave 2 used broader multiple-response checklists.
v. Top-two dichotomisation reduced the information contained in the original ordinal responses.
vi. Perceived learning enhancement is not equivalent to objectively demonstrated learning outcomes.
vii. The study was conducted at one engineering-focused TNE institution, and the Wave 2 cohort was predominantly male.
viii. Programme subgroups were small, confidence intervals were wide, and the Wave 2 Year 4 group contained only two students.
ix. English proficiency was not measured, and no non-TNE comparison group was available.
x. Exact invitation denominators were not retained, creating uncertainty about response rates and possible self-selection, and self-selection pressures may have differed between the two waves.
xi. The recovered Spring 2025 respondent-level records were used as the authoritative Wave 2 source for programme and year characteristics, directly comparable outcomes, concern profiles, desired-support profiles, and the primary adjusted model. Six outcome aggregates reported in the earlier Wave 2 article could not be reproduced from the preserved respondent-level records. Recalculation changed several descriptive values and effect estimates but did not alter the substantive conclusions. The archived export differs slightly in composition and was retained only for provenance documentation and a materially identical robustness re-estimation.

7.3. Future Research

Future studies should use identical or measurement-invariant concern scales across waves, combine repeated cross-sectional and panel components, record actual usage frequency and purposes, and link perceptions to learning and assessment outcomes. Multi-institutional studies should compare EMI-TNE, domestic Chinese, and UK contexts while measuring English proficiency directly. Pre-specified models and fully versioned, reproducible data-cleaning records are also needed to distinguish substantive changes from analytic artefacts.

Author Contributions: Conceptualisation, K.K. and D.W.; methodology, K.K., D.W. and W.S.; software, K.K., D.W. and W.S.; validation, K.K.; formal analysis, K.K.; investigation, K.K., D.W. and W.S.; resources, K.K., D.W. and W.S.; data curation, K.K., D.W. and W.S.; writing—original draft preparation, K.K.; writing—review and editing, K.K., D.W. and W.S.; visualisation, K.K.; project administration, K.K., D.W. and W.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Wave 1 was conducted as a pilot study reviewed under the ethical review processes of the Faculty of Engineering and Physical Sciences at the University of Leeds (University of Leeds Code of Research Conduct and Research Ethics; no individual reference was issued for pilot reviews). Wave 2 and the two-wave comparative design as a whole were approved by the Faculty of Engineering and Physical Sciences Research Ethics Committee, University of Leeds (Ref. 2714; 8 March 2025). Both waves were conducted in accordance with the Declaration of Helsinki.

Informed Consent Statement: Informed consent was obtained from participants in both waves. The present comparative analysis uses de-identified data and aggregate reporting.

Data Availability Statement: The de-identified respondent-level data are not publicly available because the original consent did not specify public release. Aggregate results, item wording, and analysis specifications are provided in the paper and appendices. Further access requests may be considered by the corresponding author subject to ethical and institutional approval.

Acknowledgments: The authors gratefully acknowledge Zhao Duo, Southwest Jiaotong University, for methodological comments and support for the Wave 1 pilot.

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

Appendix A. Survey Wording and Response Options

The wording below is reproduced from the labels preserved in the original Wave 1 respondent-level export and the recovered Wave 2 respondent-level export. The original Wave 1 screen layout was not retained.

Appendix A.1. Core Items

i. How familiar are you with artificial intelligence (AI)? Responses: Not at all familiar; Not very familiar; Neutral; Familiar; Very familiar.
ii. How important do you think AI will be in the field of science or engineering education in the next 5–10 years? Responses in both waves: Not important; Neutral; Important; Very important. Verified frequency distributions in that category order: Wave 1 (N = 165): 2, 21, 58, 84; recovered Wave 2 Spring cohort (N = 297): 3, 35, 102, 157.
iii. How likely would you be to use an AI chatbot to support your learning in a science or engineering programme? Responses: Very unlikely; Somewhat unlikely; Neutral; Somewhat likely; Very likely.

iv. How comfortable would you feel using an AI chatbot in a science or engineering learning context? Responses: Very uncomfortable; Somewhat uncomfortable; Neutral; Somewhat comfortable; Very comfortable.
v. How much could an AI chatbot enhance: (a) overall learning, (b) problem-solving skills, (c) collaboration skills, and (d) creativity in learning? Responses: Not at all; Slightly; Moderately; Very much; Extremely.

Appendix A.2. Wave 1 Concern and Feature Questions

i. Concern prompt (single choice): select the most important concern about using an AI chatbot in science or engineering education. Options preserved in the export: lack of human interaction; lack of personalised feedback; inability to ask follow-up questions; difficulty understanding chatbot responses; other (free text).
ii. Feature prompt (single choice): select the most desired chatbot feature. Options preserved in the export: personalised responses; ability to ask follow-up questions; option to choose chatbot personality; use for assignments or assessments; other (free text).

Appendix A.3. Wave 2 Concern Checklist

Question: "What concerns do you have about using an AI chatbot to support your learning? (Choose all that apply.)"

i. I might rely too much on the chatbot and not learn independently.
ii. I might rely too much on the chatbot and not learn independently [second exported label variant; no terminal full stop].
iii. The chatbot might give incorrect or unclear answers.
iv. The chatbot might not understand my questions clearly.
v. Lack of human interaction.
vi. Lack of personalised feedback.
vii. Concerns about academic integrity or plagiarism issues.
viii. The chatbot may struggle with complicated or advanced topics.
ix. Inability to ask follow-up questions.
x. Chatbot answers might lack depth or detailed explanations.
xi. Difficulty understanding the chatbot's responses.
xii. Fear of misunderstanding caused by translations.
xiii. It could reduce my motivation or interest in learning.
xiv. Technical issues or internet problems could limit my learning.
xv. Language barriers or difficulties communicating clearly in English.

The recovered Wave 2 respondent-level export contains 15 distinct fixed response-label strings representing 14 conceptual concerns. The over-reliance wording appears in two variants that differ only by a terminal full stop; because 51 respondents selected both variants, they were combined at the respondent level for prevalence reporting (see the note to Table 6).

Appendix A.4. Wave 2 Desired-Feature Checklist

Question: "What kinds of features or functions would you like to see in an AI chatbot that helps people learn? (Choose all that apply.)"

i. Help with understanding technical words or language.
ii. Ability to ask follow-up questions.
iii. Option to personalise responses.
iv. Step-by-step problem-solving guides.
v. Ability to translate or simplify explanations into easier language.

vi. Links to extra study resources.
vii. Option to choose a chatbot personality.
viii. Option to use the chatbot for assignments or assessments.
ix. Interactive visuals like animations or diagrams.
x. Feature to highlight or explain complex English vocabulary.
xi. Ability to ask questions or get responses in Chinese and English.
xii. Ability to answer spoken (voice) questions.
xiii. Opportunities for group or team interaction.
xiv. I have no specific preferences.

An optional "Other" text response was available in the Wave 2 concern and desired-feature blocks. Six respondents used the concern "Other" field and two used the desired-feature "Other" field. These entries were not treated as additional fixed response categories.

Appendix B. Adjusted Willingness Models

The original Wave 1 respondent-level file was recovered and used to reconstruct the complete programme × year × willingness distribution and to re-estimate the adjusted models directly at respondent level. Tables A1–A3 report the Wave 1 joint distribution, main-effects model, and programme-specific wave odds ratios from the exploratory wave × programme model, respectively.

Table A1. Wave 1 programme × year × willingness cross-tabulation.

Programme	Year 1	Year 2	Year 3	Year 4	Total
CET	8/12 (66.7%)	15/22 (68.2%)	22/28 (78.6%)	3/6 (50.0%)	48/68 (70.6%)
EEE	6/19 (31.6%)	3/4 (75.0%)	1/4 (25.0%)	6/8 (75.0%)	16/35 (45.7%)
CS	9/20 (45.0%)	4/5 (80.0%)	3/3 (100.0%)	3/4 (75.0%)	19/32 (59.4%)
ME	5/24 (20.8%)	2/2 (100.0%)	2/3 (66.7%)	0/1 (0.0%)	9/30 (30.0%)
Total	28/75 (37.3%)	24/33 (72.7%)	28/38 (73.7%)	12/19 (63.2%)	92/165 (55.8%)

Note. Each cell reports the number selecting "Somewhat likely" or "Very likely" divided by the total number in that programme–year cell, followed by the percentage. Responses of "Somewhat likely" or "Very likely" were coded as 1, while "Neutral", "Somewhat unlikely", and "Very unlikely" were coded as 0 for the binary model. Several cells are very small; this table is provided for analytical transparency rather than subgroup inference.

Table A2. Respondent-level main-effects logistic regression predicting willingness (N = 426).

Predictor	OR	95% CI	p
Wave 2025 vs. 2023	3.51	2.17–5.69	<0.001
CET vs. EEE	1.71	0.93–3.13	0.082
CS vs. EEE	1.68	0.91–3.09	0.097
ME vs. EEE	0.90	0.46–1.78	0.766
Year 2 vs. Year 1	2.21	1.11–4.40	0.024
Year 3 vs. Year 1	1.12	0.65–1.92	0.679
Year 4 vs. Year 1	1.51	0.56–4.09	0.416

Note. Reference categories are Wave 2023, EEE, and Year 1. Model fit: AIC = 499.92; Nagelkerke R^2 = 0.122.

Table A3. Exploratory wave × programme model: programme-specific Wave 2025 vs. Wave 2023 odds ratios (analytic N = 426).

Programme	Wave OR	95% CI	Interpretation
CET	1.25	0.51–3.06	No clear adjusted wave difference
EEE	4.07	1.74–9.49	Higher willingness in Wave 2

Table A3. *Cont.*

Programme	Wave OR	95% CI	Interpretation
CS	3.74	1.40–10.00	Higher willingness in Wave 2
ME	10.33	3.29–32.42	Higher willingness in Wave 2; substantial uncertainty

Interaction-model fit: AIC = 497.09; Nagelkerke R^2 = 0.149; likelihood-ratio $\chi^2(3) = 8.83$, $p = 0.032$. This model is exploratory and should not be interpreted as a programme-stage classification.

Appendix C. Multiple-Testing Families

Table A4 summarises the three Benjamini–Hochberg correction families and their key corrected results.

Table A4. Benjamini–Hochberg families and corrected values.

Family	Tests Included	Key Corrected Result
Readiness/importance	Familiarity, importance, willingness, comfort	Familiarity OR = 3.43 and willingness OR = 2.78, q < 0.001; comfort OR = 1.66, q = 0.020; importance OR = 1.10, q = 0.728
Enhancement	Overall learning, problem-solving, collaboration, creativity	Overall learning OR = 2.61, q < 0.001; problem-solving OR = 1.48, q = 0.092; collaboration OR = 1.02, q = 0.908; creativity OR = 0.91, q = 0.830
Exploratory subgroups	Four programme and four year comparisons	Eight-test family; Year 4 included in correction but not interpreted

Appendix D. Relationship to Previous Publications

The earlier publications used overlapping 2025 data but addressed distinct research questions. Neither reported the 2023 baseline or the 2023–2025 temporal comparisons. Table A5 distinguishes the previously published components from the new contributions of the present study.

Table A5. Previously published and new contributions.

Component	Status Before This Paper	Role in This Paper
Published single-site Wave 2 cohort (N = 297)	Reported item-level descriptive findings in Kajan et al. (2025)	Historical publication and provenance comparison; where aggregate values differ, the recovered Spring respondent-level records are authoritative for the present analysis
Broader 2025 multi-institution dataset (N = 467; Institution 1: n = 300)	Used in Kajan et al. (2026b) to examine lecturer-role expectations, social ambivalence, and ambivalence literacy	Used only for provenance documentation and the archived-subset robustness re-estimation; not used as the source of the primary checklist profiles
Integration-preference analysis of the broader 2025 dataset	Presented in Kajan et al. (2026a)	Distinct research question; not re-analysed here
Wave 1 baseline (N = 165)	Previously unreported	New descriptive evidence
Direct 2023–2025 comparisons	New	Primary empirical contribution, calculated using the recovered Spring respondent-level records

Table A5. *Cont.*

Component	Status Before This Paper	Role in This Paper
Adjusted wave association and sensitivity analyses	New	Primary adjusted model with an archived robustness re-estimation
Concern-maturation proposition	New theoretical proposition	Explicitly hypothesis-generating, not established empirically

References

Abdallah, N., Katmah, R., Khalaf, K., & Jelinek, H. F. (2025). Systematic review of ChatGPT in higher education: Navigating impact on learning, wellbeing, and collaboration. *Social Sciences & Humanities Open*, *12*, 101866. [CrossRef]

Albadarin, Y., Saqr, M., Pope, N., & Tukiainen, M. (2024). A systematic literature review of empirical research on ChatGPT in education. *Discover Education*, *3*, 60. [CrossRef]

Attewell, S. (2025). *Student perceptions of AI 2025*. Jisc, National Centre for AI. Available online: https://nationalcentreforai.jiscinvolve.org/wp/2025/05/21/student-perceptions-of-ai-2025/ (accessed on 10 January 2026).

Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. (2024). Developing evaluative judgement for a time of generative artificial intelligence. *Assessment & Evaluation in Higher Education*, *49*, 893–905. [CrossRef]

Bravo, F. A., & Cruz-Bohorquez, J. M. (2024). Engineering education in the age of AI: Analysis of the impact of chatbots on learning in engineering. *Education Sciences*, *14*, 484. [CrossRef]

Cabrera, C., & Neville, R. (2025). Widely used but barely trusted: Understanding student perceptions on the use of generative AI in higher education. *Perspectives: Policy and Practice in Higher Education*, 1–14. [CrossRef]

Chee, H., Ahn, S., & Lee, J. (2025). A competency framework for AI literacy: Variations by different learner groups and an implied learning pathway. *British Journal of Educational Technology*, *56*, 2146–2182. [CrossRef]

Chiu, T. K. F., Ahmad, Z., Ismailov, M., & Sanusi, I. T. (2024). What are artificial intelligence literacy and competency? A comprehensive framework to support them. *Computers and Education Open*, *6*, 100171. [CrossRef]

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*, *227*, 105224. [CrossRef]

El Fathi, T., Saad, A., Larhzil, H., Lamri, D., & Al Ibrahmi, E. M. (2025). Integrating generative AI into STEM education: Enhancing conceptual understanding, addressing misconceptions, and assessing student acceptance. *Disciplinary and Interdisciplinary Science Education Research*, *7*, 6. [CrossRef]

Farhat, Z. (2025). Hooked on help: Student overreliance on ChatGPT in higher education. *Journal of Information Technology Teaching Cases*. Advance online publication. [CrossRef]

Freeman, J. (2025). *Student generative AI survey 2025*. Higher Education Policy Institute. Available online: https://www.hepi.ac.uk/reports/student-generative-ai-survey-2025/ (accessed on 2 March 2026).

Heung, Y. M. E., & Chiu, T. K. F. (2025). How ChatGPT impacts student engagement from a systematic review and meta-analysis study. *Computers and Education: Artificial Intelligence*, *8*, 100361. [CrossRef]

Kajan, K., Shi, W., & Wanatowski, D. (2025). STEM undergraduates' perceptions of AI chatbots: A cross-sectional descriptive survey. *AI in Education*, *1*(1), 4. [CrossRef]

Kajan, K., Shi, W., & Wanatowski, D. (2026a, May 22–24). *How should AI chatbots be used in transnational STEM education? Student integration preferences and anticipated impacts on study habits* [Conference presentation]. 4th UK–China Teaching, Learning, and Educational Technology Conference (TILE-TEC 2026), Chengdu, China.

Kajan, K., Shi, W., Wanatowski, D., & Ryan, M. (2026b). Navigating the dual-view phenomenon: Social ambivalence, ambivalence literacy, and lecturer role transformation in AI-integrated transnational STEM education. *Education Sciences*, *16*, 554. [CrossRef]

Khlaif, Z. N., Alkouk, W. A., Salama, N., & Abu Eideh, B. (2025). Redesigning assessments for AI-enhanced learning: A framework for educators in the generative AI era. *Education Sciences*, *15*, 174. [CrossRef]

Kofinas, A. K., Tsay, C. H.-H., & Pike, D. (2025). The impact of generative AI on academic integrity of authentic assessments within a higher education context. *British Journal of Educational Technology*, *56*, 2522–2549. [CrossRef]

Liu, C., Zhu, K., Zhang, Z., & Ding, Y. (2025). The impact of Chinese undergraduates' perceived acceptance of AI technology on academic achievement: The mediating role of critical thinking. *Frontiers in Psychology*, *16*, 1727037. [CrossRef] [PubMed]

Liu, Z., Zuo, H., & Lu, Y. (2025). The impact of ChatGPT on students' academic achievement: A meta-analysis. *Journal of Computer Assisted Learning*, *41*, e70096. [CrossRef]

Ma, B., Li, H., Li, G., Chen, L., Tang, C., Shimada, A., & Konomi, S. (2026). Three years with classroom AI in introductory programming: Shifts in student awareness, interaction, and performance. *arXiv*. [CrossRef]

Ortiz-Bonnin, S., & Blahopoulou, J. (2025). Chat or cheat? Academic dishonesty, risk perceptions, and ChatGPT usage in higher education students. *Social Psychology of Education*, *28*, 113. [CrossRef]

Park, H., & Ahn, D. (2024). The promise and peril of ChatGPT in higher education: Opportunities, challenges, and design implications. In *Proceedings of the 2024 CHI conference on human factors in computing systems* (Article 271, pp. 1–21). Association for Computing Machinery. [CrossRef]

Parker, L., Loper, A. J., Carter, C. W., Hayes, J., & Karakas, A. (2026). Longitudinal insights into AI in education: Usage, ethics, and policy development in higher education. *Computers and Education Open*, *10*, 100329. [CrossRef]

Polyportis, A. (2024). A longitudinal study on AI adoption: Understanding drivers of ChatGPT usage behaviour change in higher education. *Frontiers in Artificial Intelligence*, *6*, 1324398. [CrossRef] [PubMed]

Ravšelj, D., Keržič, D., Tomaževič, N., Umek, L., Brezovar, N., Iahad, N. A., Abdulla, A. A., Akopyan, A., Aldana Segura, M. W., AlHumaid, J., Allam, M. F., Alló, M., Andoh, R. P. K., Andronic, O., Arthur, Y. D., Aydın, F., Badran, A., Balbontín-Alvarado, R., Ben Saad, H., . . . Aristovnik, A. (2025). Higher education students' perceptions of ChatGPT: A global study of early reactions. *PLoS ONE*, *20*(2), e0315011. [CrossRef] [PubMed]

Reuters. (2023). *ChatGPT frenzy sweeps China as firms scramble for home-grown options*. CNBC. Available online: https://www.cnbc.com/2023/02/11/chatgpt-frenzy-sweeps-china-as-firms-scramble-for-homegrown-options.html (accessed on 25 August 2025).

Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). Free Press.

Schei, O. M., Møgelvang, A., & Ludvigsen, K. (2024). Perceptions and use of AI chatbots among students in higher education: A scoping review of empirical studies. *Education Sciences*, *14*, 922. [CrossRef]

Thomson, S. R., Pickard-Jones, B. A., Baines, S., & Otermans, P. C. J. (2024). The impact of AI on education and careers: What do students think? *Frontiers in Artificial Intelligence*, *7*, 1457299. [CrossRef] [PubMed]

Wang, L., & Li, W. (2024). The impact of AI usage on university students' willingness for autonomous learning. *Behavioral Sciences*, *14*, 956. [CrossRef] [PubMed]

Wu, C., & Carroll, J. M. (2026). Evolving perspectives on generative AI in higher education: Insights from two years of university stakeholder analysis. *TechTrends*, *70*, 346–363. [CrossRef]

Xinhua. (2025). *China's generative AI users double to 515 mln: Report*. The State Council of the People's Republic of China. Available online: https://english.www.gov.cn/archive/statistics/202510/18/content_WS68f30556c6d00ca5f9a06e05.html (accessed on 10 February 2026).

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.