EUROPEAN JOURNAL OF STEM EDUCATION

European Journal of STEM Education,
2026, 11(1), 17
eISSN: 2468-4368

lectito

Research paper

Gender and Functional Differentiation in Generative AI Usage Among Malaysian Higher Education Student

Fahd Ali Raza [1*], Abtar Darshan Singh [1], Rabia Anwar [1], Jonathan Jeevan Strivinas Kovilpillai [1], Analisa Binti Hamdan [1], Fumiko Konno [1], Vaikunthan Rajaratnam [2], Murali Raman [1], Husna Hafiza Razami [1]

[1] *Asia Pacific University of Technology & Innovation,* MALAYSIA
[2] *Khoo Puat Teck Hospital,* SINGAPORE

*__Corresponding Author:__ fahd.raza@apu.edu.my

Citation: Raza, F. A., Singh, A. D., Anwar, R., Kovilpillai, J. J. S., Hamdan, A. B., Konno, F., Rajaratnam, V., Raman, M., & Razami, H. H. (2026). Gender and functional differentiation in generative AI usage among Malaysian higher education student. *European Journal of STEM Education, 11*(1), 17. https://doi.org/10.20897/ejsteme/18267

Published: March 30, 2026

ABSTRACT

Despite achieving gender parity in higher education enrolment (60% female), Malaysia faces an emerging digital divide in how students use artificial intelligence. This study examined whether gender predicts task-specific ChatGPT usage patterns among Malaysian students ($n = 443$), employing latent profile analysis and multinomial logistic regression on CC BY 4.0 licensed Global ChatGPT Student Survey data (October 2024–February 2025). Four distinct usage profiles emerged: Selective Users (14.9%), Moderate Adopters (31.8%, coding-focused), Academic Enthusiasts (33.0%, text-focused), and Comprehensive Users (20.3%). Gender significantly predicted specialized profile membership ($\chi^2 = 19.47$, $p < .001$). Males concentrated in coding-focused use ($OR = 0.48$ for females, $p = .007$), females in text-focused use ($OR = 1.89$, $p = .020$), while Comprehensive Users exhibited gender parity. Exploratory analyses indicate a tentative pattern of larger gender gaps in technical AI use in STEM fields (33.4 percentage points in Applied Sciences) versus Social Sciences (12.7 points), though small cell sizes ($n=3$ for female Natural Sciences students) prevent definitive conclusions. Profiles predicted domain-specific skill development with large effects ($\eta^2 = .18$-$.33$). Findings reveal that equal access masks unequal functional engagement, with implications for gender-segregated occupational pathways that warrant further investigation with larger samples.

Keywords: ChatGPT, higher education, gender divide, latent profile analysis, Malaysia

The November 2022 release of ChatGPT marked a defining moment in higher education, reaching 100 million users within two months, the fastest adoption rate in technological history (Gill et al., 2024; Halaweh, 2023; Rahman & Watanobe, 2023). By 2024, 89% of higher education students globally were aware of ChatGPT, with 40–60% actively using it (Xing et al., 2025). Our Malaysian subset ($n = 443$, October 2024–February 2025) shows near-universal familiarity. Yet, most research measures frequency rather than function, overlooking what students actually do with AI tools (Lin, 2023). This distinction matters: a student using ChatGPT for proofreading develops different competencies from one using it for coding, though both appear identical in adoption statistics (Lau & Guo, 2023; Lin, 2023; Xing et al., 2025).

While Malaysia has achieved gender parity in education (Raza et al., 2025; Raza & Singh, 2024), women make up 60% of higher education enrolment (Cheok, 2024) and internet penetration exceeds 89% (Hassan et al., 2021), new gendered patterns of use have emerged. Studies across Spain, Norway, and the U.S. show men using ChatGPT more for technical applications, while women favour communicative or creative uses (Møgelvang et al., 2024; Skjuve et al., 2024). These patterns are masked by traditional unidimensional adoption models such as the Technology Adoption Model (TAM) (FakhrHosseini et al., 2024; Yadegari et al., 2024). Malaysia's context heightens this concern: despite women's dominance in higher education, they occupy only 30% of ICT roles and 22–28% of engineering or computing majors (Lagesen, 2008; Saadat & Sultana, 2023; Shaari et al., 2022). National AI literacy policies (Dyczek, 2025; Low et al., 2025) assume equal access ensures equal capability, a claim that remains empirically untested.

This study investigates whether such assumptions hold by examining functional differentiation in ChatGPT use among Malaysian students. Specifically, it addresses four research questions:

RQ1 : What distinct latent profiles of task-based ChatGPT usage exist among Malaysian students?
H1 : Students will exhibit 3–5 distinct profiles representing different functional orientations.
RQ2 : To what extent does gender predict profile membership?
H2a : Gender will significantly predict profile membership ($p < .05$) after controlling for covariates.
H2b : Males will show higher probability of technically oriented profile membership.
H2c : Females will show higher probability of text-focused profile membership.
RQ3 : How do field of study, learning modality, and institution type moderate the gender–profile relationship?
H3a : Field of study will moderate gender effects (Gender × Field interaction, $p < .05$).
H3b : Gender gaps in technical profile membership will differ by field of study, with exploratory investigation of whether STEM contexts amplify or attenuate such gaps.
RQ4 : Do profiles differentially associate with perceived learning outcomes?
H4 : Technical profile students will report higher programming skill development; text-focused profile students higher writing skill development.

This study extends TAM by treating "use" as multidimensional and person-centred, offering the first systematic evidence of functional stratification in generative AI use in Southeast Asia. The findings have implications for AI literacy and curriculum design in Malaysia, emphasizing the need for interventions that promote cross-domain engagement rather than generic digital training.

LITERATURE REVIEW AND THEORETICAL FRAMEWORK

Generative AI in higher education

Research on AI in education has evolved through three technological phases. Early systems like SCHOLAR and AutoTutor were rule-based and limited to narrow domains (Wölfel et al., 2024). Second-generation platforms like ALEKS and Khan Academy used machine learning for adaptive learning but still worked within structured problem spaces (Rizvi et al., 2025). The emergence of ChatGPT in November 2022 marked a third-generation paradigm, generative, multimodal, and accessible without technical barriers (Kasneci et al., 2023; Tlili et al., 2023). Within two years, over half of university students globally reported using it (Xing et al., 2025), signalling a grassroots diffusion that preceded institutional readiness (Lalumera, 2024).

Empirical outcomes are mixed. Some studies show benefits like improved writing quality (Dergaa et al., 2023; Yan & Liu, 2025), better conceptual reasoning in STEM (Kasneci et al., 2023), and enhanced accessibility for neurodiverse learners (Biswas, 2023). Yet, concerns persist. Skill degradation, "competence traps" (Bastani et al., 2024), and reductions in lexical diversity (Huynh, 2024) point to dependency risks. Moreover, advantages concentrate among high-performing students while weaker learners show little gain (Alier et al., 2024), widening the digital learning gap. Reports of increased plagiarism further complicate institutional acceptance (Lancaster, 2023).

Across these findings runs a shared limitation: most studies dichotomize students as *users* or *non-users*, obscuring the purposes and cognitive orientations behind their engagement. This study addresses conceptual blind spots by analysing *task-based functional differentiation* through latent profile analysis, a method suited to capturing the heterogeneity of generative AI use.

THEORETICAL FOUNDATIONS

Technology acceptance model

The Technology Acceptance Model (TAM) explains technology adoption as a function of perceived usefulness (PU) and perceived ease of use (PEOU) (Davis, 1989). Later iterations, TAM2 and TAM3, integrated social influence, cognitive processes, and affective variables, increasing predictive accuracy (Venkatesh et al., 2002). Gender moderates these pathways: PEOU tends to drive men's adoption, while PU and subjective norms influence women's (Kaur & Kaur, 2022).

However, TAM assumes a unidimensional notion of "use," an assumption ill-suited to tools like ChatGPT that span technical, analytical, and creative domains (Al-Adwan et al., 2023). A learner may find ChatGPT highly useful for translation yet irrelevant for mathematical reasoning.

Digital gender gap framework

Digital inequality theory posits sequential divides: Access (Level 1), skills (Level 2), and differentiated outcomes (Level 3) (Van Deursen et al., 2021; Van Deursen & Van Dijk, 2023). In Malaysia, the first two levels have largely closed, broadband penetration exceeds 90% and digital literacy scores show no gender gaps (Hendrian, 2025; Mastam et al., 2024; Subramaniam, 2023). Yet, the third-level divide persists worldwide: women and men use identical technologies for different purposes (Işıklı & Fazlıoğlu, 2026; Lamberti et al., 2023; Raza et al., 2024; Walshe et al., 2025). Studies consistently find men concentrating on capital-enhancing or technical activities while women engage more in communicative or care-oriented uses (Awoniyi & Jokotagba, 2025; Lamberti et al., 2023; Raza et al., 2025; Sintas et al., 2023).

Several mechanisms explain these patterned behaviours: Stereotype threat (Ahn et al., 2022), disciplinary socialization (Alam, 2022), self-efficacy differences (Low et al., 2025), and opportunity structures that reinforce exposure disparities (Rahman & Halim, 2022; Raza & Singh, 2024). Together, these mechanisms highlight that equality in access does not ensure equality in outcomes, an assumption this paper empirically tests within Malaysia's higher education sector.

Task-based technology use models

Recent frameworks conceptualize technology engagement as functionally differentiated (Hofer & Hargittai, 2024; Yang & Zhang, 2023). Empirical work distinguishes reproductive (e.g., summarizing), generative (e.g., creative writing), analytical (e.g., research assistance), and technical (e.g., coding) uses, each affording distinct cognitive processes. Latent Profile Analysis (LPA) provides a statistical means to identify these user typologies (Moore & Quartiroli, 2025).

EMPIRICAL EVIDENCE ON GENDERED AI USAGE

Global patterns in Western contexts

Recent empirical work from North America and Western Europe confirms that gender remains a salient variable in generative AI adoption across diverse cultural contexts, providing a crucial benchmark for interpreting Malaysian findings. Large-scale U.S. surveys demonstrate persistent awareness and engagement gaps: 62% of men versus 53% of women had heard of ChatGPT, and among those aware, men were significantly more likely to have used it (22% vs. 15%) (Basch et al., 2025; Otis et al., 2024). European patterns mirror these trends, where 19% of men and 13% of women reported generative AI use in the preceding quarter (Otis et al., 2024). Educational applications in Western universities show similar functional stratification patterns. Recent scholarship highlights how digital innovation and cloud-based technologies are transforming learning experiences in ways that intersect with structural inequalities, including gender (Papadakis et al., 2024), while longitudinal analyses of social justice in education reveal persistent equity gaps that frame the third-level digital divide examined here (Karakose et al., 2023). Lampropoulos and Papadakis (2025) demonstrate that AI and social robot integration in European educational settings reveals gendered engagement patterns, with male students showing higher technical interaction rates while female students prioritize communicative applications. This pattern emerges even in primary education: Uğraş et al. (2024) found that when ChatGPT was introduced for pedagogical support in Turkish primary schools, boys gravitated toward computational tasks while girls focused on language-based activities, suggesting early-emerging functional differentiation that precedes higher education. This primary school pattern contrasts with our higher education context in important ways. While elementary students showed functional differentiation during initial AI exposure, our Malaysian university students have already completed years of disciplinary socialization. The persistence of gendered patterns from primary school through university suggests these preferences may solidify rather than diminish with educational progression, though longitudinal studies are needed to test this developmental hypothesis.

Among higher education populations in Western universities, patterns remain consistent yet nuanced. Alier et al. (2024) observed higher overall adoption among Spanish male students (71%) than females (62%), with the largest gaps in engineering and computing disciplines. Critically, gender effects were partially mediated by self-efficacy and anxiety, underscoring the psychological underpinnings of digital participation. Strzelecki (2024) reported negligible adoption differences among Norwegian students but significantly higher self-assessed AI literacy among men, indicating that confidence gaps may operate independently of usage frequency. Similarly, U.S. studies found that while students used ChatGPT widely, males were twice as likely to engage with technical or coding tasks, and females more with language learning (Kelly et al., 2025; Shrivastava, 2024). Cross-disciplinary comparisons in Western contexts suggest the gender gap is field-contingent rather than universal. Disparities are most pronounced in engineering and computing (15-20 percentage point gaps), moderate in natural sciences, and often statistically insignificant in health, education, or social sciences (Alier et al., 2024; Strzelecki, 2024). These findings align with the broader literature on gendered occupational cultures in computing (Clarke et al., 2024; Klinger & Svensson, 2023; Li, 2023), where masculine identity norms influence both confidence and tool orientation.

Cross-cultural comparison: Western vs. Malaysian contexts

The convergence of findings across Western and Southeast Asian contexts suggests that gendered functional differentiation in AI use represents a global phenomenon rather than a culture-specific artifact. However, three notable divergences merit attention and frame our Malaysian investigation. Firstly, magnitude of technical gaps. Western studies report coding-focused usage gaps of 15-20 percentage points between males and females in STEM fields (Alier et al., 2024; Strzelecki, 2024). Our Malaysian data reveals substantially larger gaps, 33.4 percentage points in Applied Sciences, suggesting potential cultural or institutional amplification. This difference may reflect Malaysia's more pronounced disciplinary gender segregation, where women comprise only 22-28% of engineering/computing students compared to 35-40% in European programs (Saadat & Sultana, 2023). The question driving our analysis is whether Malaysia's educational gender parity (60% female enrolment) coexists with intensified functional stratification.

Secondly, comprehensive engagement rates. Norwegian and Spanish samples show 25-30% comprehensive AI usage rates (Strzelecki, 2024), compared to preliminary indications of lower rates in Malaysian contexts. This difference likely stems from earlier AI literacy integration in Western curricula, where ChatGPT training began in Spring 2023 compared to Fall 2024 in Malaysian universities. Our study examines whether this temporal lag affects the distribution of usage profiles and whether latecomer status intensifies or mitigates gendered patterns.

Thirdly, field moderation patterns. Western studies report smaller gender gaps in natural sciences than social sciences (Alier et al., 2024), suggesting that technical training may equalize appropriation. However, preliminary Malaysian data hints at the opposite pattern, amplified gaps in STEM contexts. This potential "tentative STEM amplification pattern" contradicts conventional diversity intervention assumptions and requires rigorous empirical testing.

There are theoretical implications of cross-cultural patterns. Gendered differentiation appears consistent across cultures. Both Western and Southeast Asian contexts show males gravitating toward technical uses and females toward communicative uses. This pattern suggests universal mechanisms like stereotype threat (Ahn et al., 2022), identity alignment (Sintas et al., 2023; Webb, 2026), and self-efficacy differences (Low et al., 2025). However, magnitude differences point to culturally specific moderators including disciplinary segregation intensity, intervention timing, and institutional support structures. Malaysia's unique positioning, middle-income, Muslim-majority, multilingual, digitally advanced, with educational gender parity yet occupational segmentation, enables critical tests of whether third-level digital divides persist across cultural boundaries or whether specific sociocultural configurations intensify functional stratification. Our person-centred latent profile approach moves beyond simple gap detection to identify qualitatively distinct usage patterns, enabling more nuanced cross-cultural comparison than variable-centred methods employed in Western studies.

Synthesis and study positioning

Collectively, this body of evidence from Western and Malaysian contexts points to consistent yet undertheorized gender differences in generative AI engagement, differences often masked by aggregate adoption measures. Existing studies tend to rely on frequency scales and variable-centred analyses, limiting insight into how and why gendered task differentiation occurs. The present study addresses these shortcomings through three methodological advances. First, we adopt a person-centred latent profile approach that identifies qualitatively distinct usage patterns rather than assuming homogeneous effects across students. This enables more nuanced cross-cultural comparison than the variable-centred methods employed in Western studies. Second, we integrate TAM and digital divide frameworks to examine both individual-level predictors (gender, field) and contextual moderators (institution type, learning format). Third, we situate our analysis in a non-Western context where educational gender parity (60% female enrolment) paradoxically coexists with occupational segregation (women comprise 22-28% of engineering/computing majors), providing a critical test of whether functional stratification operates universally or intensifies in specific sociocultural configurations. Our four research questions systematically build on Western findings while testing their generalizability: (RQ1) Do Malaysian students exhibit similar profile structures to those implied by Western task-frequency data? (RQ2) Does gender predict specialized profile membership as Western studies suggest? (RQ3) Do STEM contexts amplify or reduce gender gaps, testing the "technical training equalization" assumption? (RQ4) Do profiles predict domain-specific outcomes, validating functional distinctions?

Malaysian context

Malaysia offers a distinctive setting to examine gendered AI engagement due to its juxtaposition of educational parity and occupational segmentation. The national higher education system encompasses 20 public and 47 private universities serving over 1.2 million students (Wickneswary et al., 2024). Women represent the majority at every level, 61% of undergraduates, 58% of master's, and 52% of doctoral enrolments, yet remain concentrated in feminized disciplines such as education, health sciences, and social sciences, while comprising only 22–28% of students in engineering and computer science (Elhadary & Samat, 2023; Saadat & Sultana, 2023; Zhou, 2021).

Technological infrastructure is robust, with broadband penetration exceeding 90% in urban and 89% in rural areas, and smartphone ownership surpassing 90% among young adults (Hassan, 2024; Nohuddin et al., 2025). Policy initiatives such as the National 4IR Policy (Jamaluddin et al., 2025; Taib et al., 2025) and the Malaysia Digital Economy Blueprint (Rahim & Iqbal, 2025) mandate AI literacy integration across educational levels. Post-pandemic reforms have normalized blended learning in 67% of universities (Mustapha et al., 2022; Seong et al., 2022; Soon Tan et al., 2022), creating conditions conducive to large-scale generative AI use. However, the assumption that infrastructural readiness translates into equitable digital capability remains largely untested.

Sociocultural dynamics further shape technology use. Malaysia's collectivist orientation encourages conformity and peer modelling (Raza et al., 2024; Raza & Singh, 2024), while its Islamic ethical frameworks and multilingual learning environments add layers of interpretive complexity (Sharipova, 2022; Kuah et al., 2021; Patras et al., 2022). Prior research indicates minimal gender difference in general computer self-efficacy but persistent disparities in programming confidence (Elhadary & Samat, 2023; Saadat & Sultana, 2023; Zhou, 2021) and a stronger influence of subjective norm on women's technology acceptance (Rahman & Halim, 2022; Zhou, 2021).

Few studies have examined ChatGPT adoption in Malaysian higher education, highlighting the need to explore how gender, discipline, and institutional context shape generative AI use. This study addresses that gap using a person-centred latent profile approach that integrates the Technology Acceptance Model and digital divide frameworks to identify usage patterns and their predictors.

METHODOLOGY

Research design, data source, and sample

Guided by TAM and third-level digital divide frameworks, this study adopted a cross-sectional, person-centred design to examine functional differentiation in ChatGPT use among Malaysian higher education students. Secondary data were drawn from the Global ChatGPT Student Survey conducted by the University of Ljubljana with over 200 international partners (Aristovnik et al., 2025). The global dataset (Mendeley Data, https://doi.org/10.17632/nv2343nwsb.2) contains 22,963 valid responses from 120 countries collected between October 2024 and February 2025.

Analyses focused on Malaysian respondents ($n = 443$; 1.9% of the global sample) who met three inclusion criteria: (1) enrolment at a Malaysian higher education institution, (2) age ≥ 18 years, and (3) prior ChatGPT use. The study aimed to: identify latent task-based usage profiles, test gender as a predictor of profile membership, examine disciplinary moderation, and validate profiles through perceived learning outcomes.

The parent study employed convenience sampling through instructor referrals. Participants completed the English-language survey via the 1KA web platform (https://www.1ka.si/d/en) after providing digital informed consent. The analytical sample comprised 443 Malaysian students (58.0% female, 40.9% male, 1.1% other/prefer not to say). Five non-binary/undisclosed cases ($n < 10$) were excluded from gender-stratified analyses to ensure statistical stability, yielding $n = 438$ (257 female, 181 male). Sensitivity tests confirmed negligible impact (Δ parameters < .02).

Respondents' mean age was 22.4 years ($SD = 3.8$). Field distribution: Social Sciences 48.3%, Applied Sciences 33.6%, Arts & Humanities 11.5%, Natural Sciences 6.5%. Most studied at public institutions (67.3%), primarily undergraduates (89.4%), and engaged mainly in blended learning (58.9%). A significant Gender × Field association emerged ($\chi^2 = 17.85$, $df = 3$, $p < .001$): women were over-represented in social and humanistic disciplines and under-represented in applied and natural sciences. **Table 1a** summarizes sample characteristics.

Table 1a

Sample demographics and ChatGPT usage characteristics (n = 443)

Characteristic	Category	n	%
Gender	Female	257	58.0
	Male	181	40.9
	Other/PNTS	5	1.1
Field of Study	Social Sciences	214	48.3
	Applied Sciences	149	33.6
	Arts & Humanities	51	11.5
	Natural Sciences	29	6.5
Academic Level	Undergraduate (Years 1-4)	396	89.4
	Postgraduate	47	10.6
Institution Type	Public	298	67.3
	Private	145	32.7
Learning Modality	Fully Online	49	11.1
	Blended/Hybrid	261	58.9
	Fully Face-to-Face	133	30.0
Age	18-21 years	198	44.7
	22-25 years	187	42.2
	26+ years	58	13.1
Experience Level	Beginner	89	20.1
	Intermediate	247	55.8
	Advanced	107	24.2
Usage Frequency	Rarely	71	16.0
	Sometimes	186	42.0
	Often	142	32.1
	Very Often/Always	44	9.9

Note: Mean age: 22.4 years (SD = 3.8). Mean usage intensity: 2.36 (SD = 0.87) on 1-4 scale.

Instrumentation

The parent survey underwent multi-stage validation involving expert review (n = 5), cognitive interviews (n = 8), and pilot testing (n = 45), resulting in a 62-item instrument across 11 themes (Aristovnik et al., 2025). Median completion time was 13.2 minutes.

Dependent variable: Task-specific ChatGPT use. Twelve items (Q18a–Q18l) captured usage frequency across academic and non-academic domains: writing, proofreading, brainstorming, translation, summarizing, calculation, study and research assistance, creative and professional writing, coding, and personal assistance. Responses were rated on a 5-point Likert scale (1 = Never, 5 = Always). Internal consistency was high (Cronbach's α = .89) with adequate variance (SD = 1.02–1.28) and no ceiling effects.

Primary predictor: Gender. Gender (Q2) was coded 0 = male, 1 = female; five "other/prefer not to say" responses were omitted due to power limitations. While recognizing gender's multidimensional nature, the small non-binary subsample precluded stable estimation; this exclusion is acknowledged in the limitations.

Moderator: Field of study. Disciplines were grouped into four categories based on ISCED: (1) Social Sciences (business, education, law, psychology), (2) Applied Sciences (engineering, computer science, medicine, architecture), (3) Arts & Humanities, and (4) Natural Sciences. Two independent raters coded open responses (Cohen's κ = .96) (Cohen, 2013).

Covariates. Institution type (public/private), academic level (ordinal 1–5), learning modality (dummy-coded; blended reference), age (continuous), usage intensity (ordinal 1–4), and experience (ordinal 1–3) were included to control for contextual differences.

Outcome variables. Five composites validated profile distinctions (Table 1b). All multi-item scales demonstrated strong reliability (α > .85). Single-item measures displayed sufficient variance (SD = 0.94–1.14).

Table 1b

Perceived skill development outcomes

Outcome	Example Item	α
Academic Enhancement	"ChatGPT can improve my general knowledge."	.89
Writing Skills	"ChatGPT can improve my academic writing proficiency."	.91
Programming Skills	"ChatGPT can improve my programming skills."	-
Data Analysis Skills	"ChatGPT can improve my data analysis skills."	-
Overall Satisfaction	"I am satisfied with the level of assistance provided by ChatGPT."	-

Note: All items measured on 5-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree). Single-item measures demonstrated adequate variance (SD = 0.94-1.14). The full survey instrument with all selected items, response scales, and coding notes is provided in Appendix C.

Analytical procedures

Analyses followed six integrated phases: (1) descriptive and bivariate statistics; (2) latent profile analysis (LPA); (3) multinomial logistic regression; (4) Gender × Field moderation; (5) ANOVA-based profile validation; and (6) robustness checks. All analyses were conducted in R 4.3.1 using tidyLPA, nnet, mice, emmeans, and ggplot2. Significance was set at $\alpha = .05$, emphasizing effect sizes (OR, Cramér's V, η^2) over p-values (Pérez, 2024).

Latent profile analysis

LPA identified subpopulations of students exhibiting similar mean responses across the 12 task indicators. Models with K = 2–6 profiles were estimated using full-information maximum likelihood to address 2.3% missing data. Model selection considered BIC, AIC, entropy (>.80), average posterior probabilities (>.80), BLRT ($p < .05$), interpretability, and practical profile sizes (≥10%). The final K was chosen by balancing quantitative fit with conceptual coherence (Nylund-Gibson et al., 2023).

Multinomial logistic regression and moderation

A multinomial logistic model tested whether gender predicted profile membership while controlling for covariates. The model specification was:

$$\log [P(Y_i = j \mid X_i) / P(Y_i = 4 \mid X_i)] = \beta_{0j} + \beta_{1j}X_{1i} + \Sigma_{k=2}^{K} \beta_{kj}X_{ki}$$

Where:

- $Y_i \in \{1, 2, 3, 4\}$ represents profile membership (Selective, Moderate, Academic, Comprehensive Users)
- $j \in \{1,2,3\}$ indexes non-reference categories
- $j = 4$ (Comprehensive Users) serves as the reference category
- X_{1i} = gender (0 = male, 1 = female)
- $X_{2i}...X_{ki}$ represent K-1 covariates (field, institution type, modality, age, usage intensity, experience)
- The model estimates separate coefficients (β_{0j}, β_{1j}, $\beta_{2j}...\beta_{kj}$) for each non-reference category

Estimation used multiple imputation (m = 20 datasets) with Rubin's rules for pooling across imputations (Austin et al., 2021; Blazek et al., 2021; Lydersen, 2022). The multinomial logit link function assumes the Independence of Irrelevant Alternatives (IIA); this assumption was verified using Hausman-McFadden tests (all $p > .10$, supporting IIA). Model fit was acceptable (Nagelkerke pseudo-$R^2 = .24$; likelihood ratio test $\chi^2(36) = 187.4$, $p < .001$). Variance Inflation Factors for all predictors were below 2.5, indicating no problematic multicollinearity.

Profile validation

To test construct validity, a one-way ANOVA compared perceived learning outcomes across latent profiles:

$$\text{Outcome}_{ij} = \mu + \alpha_j + \varepsilon_{ij}$$

Post-hoc Tukey HSD tests controlled for family-wise error ($\alpha = .05$). Effect sizes (η^2) were interpreted using Cohen's benchmarks (.01 = small, .06 = medium, .14 = large) (Cohen, 2013). Profiles differing significantly in outcomes supported functional interpretability.

Robustness and missing data

Robustness checks evaluated analytic stability through (a) K-means clustering (K = 4; ARI = .81), (b) 70/30 train-test cross-validation ($\chi^2 = 0.89$, $p = .83$), (c) outlier exclusion (n = 8), (d) complete-case versus imputed datasets (Δ ORs < 0.10), (e) multiple-group LPA for measurement invariance ($\Delta\chi^2(36) = 44.8$, $p = .15$), and (f) bootstrap CIs (1,000 resamples). Results across methods were consistent.

Missingness averaged 2.3% for usage items and < 1% for demographics. Little's MCAR test ($\chi^2 = 147.3$, df = 138, $p = .283$) supported the MCAR assumption. Full-information maximum likelihood was used for LPA, and multiple imputation for regression, avoiding bias from listwise deletion (Depaoli et al., 2025; Homeister, 2024; Woods et al., 2024).

Ethical considerations

The Technology Acceptance Model (TAM) explains technology adoption as a function of perceived usefulness (PU) and perceived ease of use (PEOU) (Davis, 1989). Later iterations, TAM2 and TAM3, integrated social influence, cognitive processes, and affective variables, increasing predictive accuracy (Venkatesh et al., 2002). Gender moderates these pathways: PEOU tends to drive men's adoption, while PU and subjective norms influence women's (Kaur & Kaur, 2022).

Secondary data ethics compliance

The parent dataset is published under a CC BY 4.0 license (Aristovnik et al., 2025), which permits unrestricted reuse, redistribution, and analysis for any purpose including demographic subgroup investigations. To mitigate re-identification risk in small subgroups (e.g., Female × Natural Sciences, n=3), results from this cell are reported only as aggregate interaction estimates and are explicitly flagged as statistically uninformative throughout. Ethics approval was obtained from institutional committees at participating universities in compliance with the Declaration of Helsinki. Participants consented to: (1) de-identified data being included in a public repository for research reuse, (2) secondary analyses examining relationships between demographic characteristics (including gender, field of study, institution type) and technology usage patterns, and (3) aggregated results being published in academic outlets. The consent protocol specifically disclosed that anonymised data would be made publicly available via Mendeley Data for scholarly reuse, ensuring participants were aware their responses could be analysed beyond the primary research team.

All data received by our research team were fully anonymised prior to access, with no personally identifiable information (names, emails, IP addresses, specific institutional identifiers beyond country-level aggregation) retained in the dataset. The Malaysian subset analysis (n = 443) further aggregates responses such that individual identification is statistically impossible even through indirect means. Participant anonymity was preserved through: (1) removal of all direct identifiers before public data deposit, (2) categorical rather than continuous demographic variables preventing outlier identification, (3) suppression of rare category combinations (e.g., specific programme-institution pairs with n < 5), and (4) aggregate-level reporting ensuring no individual response pattern is discernible.

Ethical approval for secondary analysis

No additional ethics approval was required for this secondary analysis as: (a) data were publicly available and de-identified at the point of access, (b) analyses fell within the scope of the parent study's consent framework, which explicitly permitted demographic subgroup investigations, (c) no contact with original participants occurred, and (d) institutional review exemption applies for secondary analysis of public datasets per Malaysian research ethics guidelines and international standards (Category 4: Benign Behavioural Interventions/Existing Data under Common Rule 45 CFR 46.104(d)(4)). The publicly archived nature of the dataset with explicit reuse permissions eliminates concerns about scope creep beyond original consent, as participants were informed their data would support multiple research projects examining educational technology use.

Ethical considerations for gender analysis

Gender was measured via self-identification with options: "Male," "Female," "Other," and "Prefer not to say." The exclusion of five participants selecting "Other" or "Prefer not to say" (1.1% of the sample) was a methodological necessity to preserve statistical validity given insufficient power for stable estimation in this subgroup (post-hoc power < 10% for any effect involving n = 5). This exclusion does not imply value judgements about gender diversity but rather reflects sample size constraints that precluded scientifically defensible inference for non-binary participants.

We acknowledge this limitation has ethical implications beyond statistical considerations. Non-binary students navigating gendered technology cultures may face unique challenges that our binary analytical framework cannot capture. Their exclusion from analysis means their experiences remain invisible in our findings, potentially reinforcing the gender binary in policy and practice despite our awareness that gender operates on a continuum. Future research with purposive sampling of non-binary students (minimum n ≥ 50) is essential to understand their technology appropriation patterns, which may differ from binary categories due to reduced gender-schema constraints or, conversely, intensified marginalisation in masculine technical spaces.

Interpretation framework

Gender was measured via self-identification with options: "Male," "Female," "Other," and "Prefer not to say." The exclusion of five participants selecting "Other" or "Prefer not to say" (1.1% of the sample) was a methodological necessity to preserve statistical validity given insufficient power for stable estimation in this subgroup (post-hoc power < 10% for any effect involving n = 5). This exclusion does not imply value judgements about gender diversity but rather reflects sample size constraints that precluded scientifically defen

Findings are interpreted as reflecting socially constructed gendered patterns of technology appropriation (Unger, 2020) rather than biological essentialism (Devitt, 2023). Gender differences observed in this study represent behavioural responses to social, cultural, and institutional contexts, including stereotype threat, peer influence, disciplinary cultures, and identity norms, not innate capabilities or preferences. Our theoretical framework explicitly positions gender as a social category that shapes but does not determine technology engagement, recognising that individuals actively negotiate, resist, or reproduce gendered expectations in their AI appropriation practices.

This social constructionist perspective has methodological implications: we expect gender effects to vary across cultural contexts, diminish with institutional intervention, and interact with other social categories (race, class, sexuality) in complex ways. Our Malaysian findings should not be interpreted as evidence of universal, fixed gender differences but rather as documenting how gender currently operates within specific educational and technological configurations that are amenable to change through policy and cultural intervention.

Limitations and strengths

Convenience sampling may favour AI-interested, English-proficient, urban students, and the cross-sectional design limits causal inference. Self-report measures introduce potential recall and desirability bias. Binary gender operationalization excludes non-binary experiences, and unmeasured mediators such as self-efficacy or stereotype threat remain outside model scope.

Nevertheless, the study provides several methodological and contextual strengths. The person-centred LPA captures heterogeneity in functional use that frequency-based analyses overlook. The 12-item measurement offers fine-grained resolution of third-level digital divide patterns. Seven covariates and interaction tests enhance model precision, and six robustness checks confirm stability. Contextually, the Malaysian setting, middle-income, Muslim-majority, multilingual, and digitally advanced, extends gender and AI literature beyond Western samples. Timing two years post-ChatGPT launch captures a critical equilibration period when AI integration had become normalized in hybrid learning environments.

These analytic procedures collectively ensured that identified latent profiles represent meaningful, replicable differences in functional engagement with generative AI. The resulting models enable rigorous testing of gendered and disciplinary patterns of ChatGPT use in Malaysian higher education.

RESULTS

Preliminary analyses

The analytical sample comprised 443 Malaysian students meeting inclusion criteria (currently enrolled, age ≥18, ChatGPT users, <25% missing data). Gender: 56.9% female, 42.0% male, 1.1% other/prefer not to say. Age ranged 18-67 years (M = 23.49, SD = 6.47, Mdn = 22), with 78.1% aged 18-24. Study level: 86.9% undergraduate, 9.9% postgraduate, 3.2% doctoral. Institution: 62.3% public, 37.7% private. Learning modality: 58.9% blended, 30.2% traditional, 10.8% fully online. ChatGPT version: 90.3% free, 2.0% paid, 7.7% both. Usage frequency: 7.9% rarely, 15.6% occasionally, 47.6% moderately, 21.0% considerably, 7.9% extensively. Experience rating: 71.7% good/very good. Descriptive statistics for task-specific ChatGPT use are presented in **Table 2**.

Task-specific usage patterns

Most frequent uses: brainstorming (M = 3.48), summarizing (M = 3.43), academic writing (M = 3.34). Least frequent: calculating (M = 2.56), coding (M = 2.61), creative writing (M = 2.67). Standard deviations (1.11-1.38) indicate substantial individual differences, with coding showing highest variability.

Correlation matrix revealed three task clusters, namely writing-related (academic writing, proofreading, summarizing) intercorrelated at r = .69-.73; analytical (study, research, summarizing) at r = .71-.74; coding correlated moderately with all (r = .38-.59), strongest with calculating (r = .59). No multicollinearity concerns (highest r = .74, all VIF < 2.5).

Table 2

Descriptive statistics for ChatGPT task usage (n = 443)

Task	M	SD	Skewness	Kurtosis
Q18a: Academic writing	3.34	1.13	-0.37	-0.72
Q18b: Professional writing	2.78	1.21	0.10	-0.98
Q18c: Creative writing	2.67	1.23	0.22	-0.96
Q18d: Proofreading	3.01	1.26	-0.06	-1.05
Q18e: Brainstorming	3.48	1.12	-0.53	-0.49
Q18f: Translating	2.98	1.31	0.01	-1.10
Q18g: Summarizing	3.43	1.11	-0.47	-0.58
Q18h: Calculating help	2.56	1.28	0.32	-0.93
Q18i: Study assistance	3.27	1.19	-0.30	-0.84
Q18j: Personal assistance	2.92	1.23	0.05	-0.96
Q18k: Research assistance	3.25	1.16	-0.32	-0.78
Q18l: Coding assistance	2.61	1.38	0.30	-1.09

Note: Scale: 1 = Never, 5 = Always. Missing data: 1.1-2.7% per item.

Gender differences in task usage

Significant differences in 4 of 12 tasks: males higher on coding ($r = .29$, largest gap) and calculating ($r = .17$); females higher on creative writing ($r = .14$) and personal assistance ($r = .10$). Core academic tasks (academic writing, proofreading, summarizing, brainstorming, study assistance, research assistance) showed no gender differences, indicating equitable engagement with mainstream educational uses. Gender differences across all 12 tasks are reported in Table 3.

Table 3

Gender differences in ChatGPT task usage (mann-whitney U tests)

Task	Male M (SD)	Female M (SD)	Z	p	r	Interpretation
Q18a: Academic writing	3.23 (1.18)	3.42 (1.09)	-1.67	.095	.08	No difference
Q18b: Professional writing	2.69 (1.24)	2.84 (1.19)	-1.23	.219	.06	No difference
Q18c: Creative writing	2.46 (1.22)	2.82 (1.23)	-2.91	.004**	.14	Female > Male
Q18d: Proofreading	2.87 (1.30)	3.11 (1.23)	-1.88	.060	.09	No difference
Q18e: Brainstorming	3.37 (1.17)	3.57 (1.08)	-1.73	.084	.08	No difference
Q18f: Translating	2.93 (1.35)	3.02 (1.29)	-0.75	.455	.04	No difference
Q18g: Summarizing	3.32 (1.17)	3.51 (1.06)	-1.68	.092	.08	No difference
Q18h: Calculating help	2.81 (1.32)	2.38 (1.24)	-3.49	<.001***	.17	Male > Female
Q18i: Study assistance	3.16 (1.24)	3.35 (1.16)	-1.43	.153	.07	No difference
Q18j: Personal assistance	2.76 (1.26)	3.03 (1.20)	-2.18	.029*	.10	Female > Male
Q18k: Research assistance	3.23 (1.20)	3.27 (1.14)	-0.49	.623	.02	No difference
Q18l: Coding assistance	3.10 (1.43)	2.26 (1.26)	-6.14	<.001***	.29	Male > Female

*Note: n = 186 males, 252 females. Effect size r: .10 = small, .30 = medium, .50 = large. *p < .05, **p < .01, ***p < .001.*

Gender × field cross-tabulation

Significant gender-field association ($\chi^2 = 17.85$, $p < .001$): females overrepresented in Arts/Humanities (17.9% vs. 8.1%, $z = +2.6$), males in Applied Sciences (39.8% vs. 29.8%, $z = +2.1$) and Natural Sciences (6.5% vs. 1.2%, $z = +2.8$). Social Sciences relatively balanced. This field segregation necessitates multivariate control. The full cross-tabulation of gender by field of study is shown in Table 4.

Table 4

Cross-tabulation of gender × field of study

Field	Male n (%)	Female n (%)	Total n (%)	Standardized Residual
Arts & Humanities	15 (8.1%)	45 (17.9%)	60 (13.7%)	Male: -2.6**, Female: +2.6**
Social Sciences	85 (45.7%)	129 (51.2%)	214 (48.9%)	Male: -1.1, Female: +1.1
Applied Sciences	74 (39.8%)	75 (29.8%)	149 (34.0%)	Male: +2.1*, Female: -2.1*
Natural Sciences	12 (6.5%)	3 (1.2%)	15 (3.4%)	Male: +2.8**, Female: -2.8**
Total	186 (100%)	252 (100%)	438 (100%)	$\chi^2(3) = 17.85$, $p < .001$

*Note: Cramér's V = .202 (small-to-medium effect). *p < .05, **p < .01.*

Latent profile analysis

Model selection

Model selection prioritized interpretability, classification quality, and practical class sizes alongside statistical fit indices. While the 5-class solution showed lower BIC (13,875.2) and AIC (13,518.2) compared to the 4-class solution (BIC = 13,941.5; AIC = 13,645.8), suggesting marginal improvement in model fit, three factors favoured the 4-class solution:

First, the Lo-Mendell-Rubin Likelihood Ratio Test (LMRT) indicated that adding a fifth class did not significantly improve fit (K=5: p = .184), whereas the fourth class provided significant improvement over three classes (K=4: p = .021).

Second, the 5-class solution produced a class containing only 7.7% of the sample (n = 34), below the 10% minimum recommended for stable estimation (Nylund-Gibson & Choi, 2018). In contrast, the 4-class solution maintained all classes ≥14.9% (minimum n = 66).

Third, the 4-class solution achieved higher entropy (.871 vs. .849), indicating superior classification certainty. All average posterior probabilities exceeded .89 (Profile 1: .927; Profile 2: .893; Profile 3: .906; Profile 4: .891), and odds of correct classification exceeded 28 for all profiles, demonstrating excellent separation.

Fourth, substantive interpretation supported four profiles. The 5-class solution divided the "Moderate Adopters" profile into two nearly identical subgroups differing only in coding intensity (M = 3.1 vs. 3.7), representing quantitative rather than qualitative distinctions.

Balancing parsimony, interpretability, and statistical criteria, the 4-class solution was retained. Sensitivity analyses using k-means clustering independently identified four clusters with high agreement (Adjusted Rand Index = .81), providing external validation of the profile structure. Fit indices for all latent profile models are summarised in **Table 5**.

Table 5

Fit indices for latent profile models (K = 2 to 6)

K	LogLik	Parameters	AIC	BIC	Entropy	LMRT p	Min Class %	Min Class n
2	-7,128.4	38	14,332.8	14,505.9	.856	<.001	31.8%	141
3	-6,891.2	51	13,884.4	14,118.8	.862	.008	20.3%	90
4	-6,758.9	64	13,645.8	13,941.5	.871	.021	14.9%	66
5	-6,682.1	77	13,518.2	13,875.2	.849	.184	7.7%	34
6	-6,631.8	90	13,443.6	13,861.9	.837	.298	3.8%	17

Note: Bold indicates selected model. All models converged successfully.

Profile characterization

Profile 1 selective users (14.9%): Uniformly low usage (all M = 1.39-1.85). Minimal adopters. Profile 2 moderate adopters (31.8%): Moderate use overall (M = 2.24-3.16) with elevated coding (M = 3.42). Coding 1.2-1.5 points higher than other tasks. Technically-oriented moderate users. Profile 3 academic enthusiasts (33.0%): High text-processing: academic writing (M = 4.18), proofreading (M = 3.88), summarizing (M = 4.21), brainstorming (M = 4.08), study (M = 4.07), research (M = 4.03). Very low coding (M = 1.98, lowest). Text-focused, avoiding technical domains.

Profile 4 comprehensive users (20.3%): High engagement across all 12 tasks (all M > 3.9). Brainstorming and summarizing near ceiling (M = 4.64). Power users exhibiting breadth.

Classification quality: Entropy = .871, all AvePP > .89 (Profile 1: .927, Profile 2: .893, Profile 3: .906, Profile 4: .891), all OCC > 28. Excellent separation.

Figure 1 displays the radar chart visualization of mean task usage frequencies across all four profiles, illustrating the distinct functional patterns: Selective Users' uniformly low engagement, Moderate Adopters' elevation on coding, Academic Enthusiasts' text-focused specialization, and Comprehensive Users' consistently high breadth.

Latent Profile Patterns - Mean Task Usage by Profile. Radar chart displays mean usage frequency (1 = Never to 5 = Always) across 12 ChatGPT task domains for four latent profiles identified via LPA. Profile 1 (Selective Users, 14.9%, grey) exhibits uniformly low engagement across all tasks (M = 1.39-1.85). Profile 2 (Moderate Adopters, 31.8%, blue) shows moderate use overall with distinctive elevation on coding assistance (M = 3.42), 1.2-1.5 points higher than other tasks. Profile 3 (Academic Enthusiasts, 33.0%, orange) demonstrates high engagement in text-processing domains (academic writing M = 4.18, proofreading M = 3.88, summarizing M = 4.21, brainstorming M = 4.08) with very low coding use (M = 1.98, lowest across all profiles). Profile 4 (Comprehensive Users, 20.3%, green) exhibits consistently high usage across all 12 domains (all M > 3.9), with brainstorming and summarizing near ceiling (M = 4.64). Profile separation validates functional differentiation: specialized patterns

(Moderate coding-focused, Academic text-focused) versus comprehensive breadth. Classification quality: entropy = .871, all AvePP > .89. Item-response means by latent profile are detailed in **Table 6**.

Figure 1

Latent profile patterns - Mean task usage byprofile

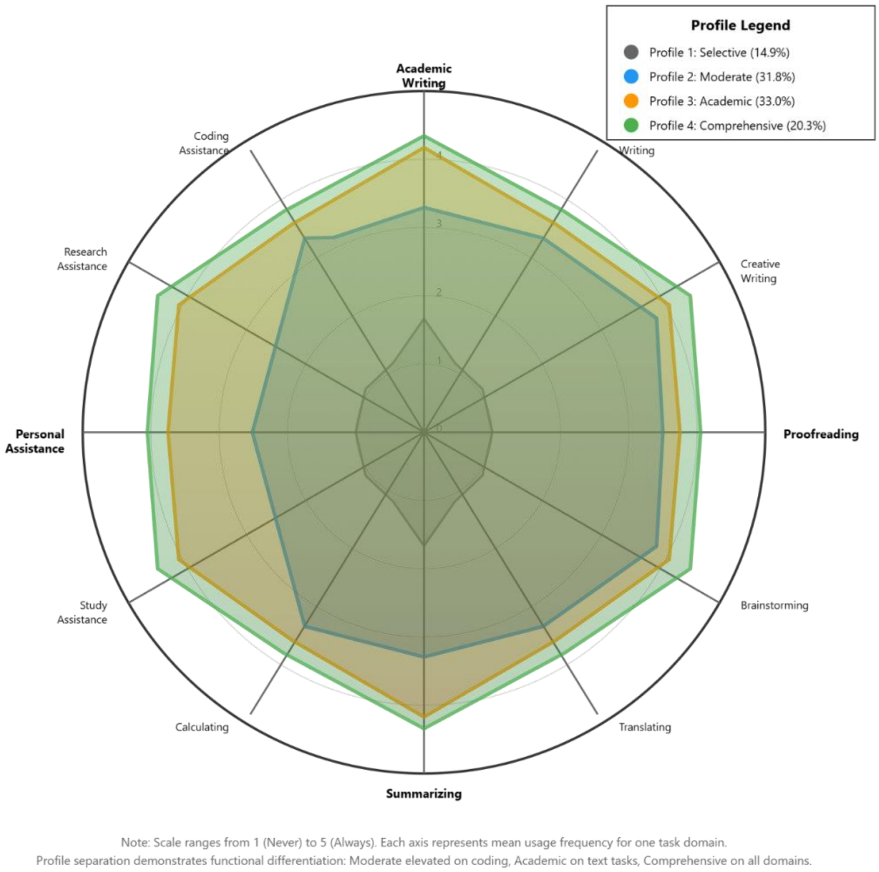


Table 6

Item-response means by latent profile (4-class solution)

Task	Overall M	Profile 1: Selective (14.9%)	Profile 2: Moderate (31.8%)	Profile 3: Academic (33.0%)	Profile 4: Comprehensive (20.3%)
Q18a: Academic writing	3.34	1.67	2.94	4.18	4.34
Q18b: Professional writing	2.78	1.42	2.24	3.26	4.03
Q18c: Creative writing	2.67	1.39	1.95	3.01	4.12
Q18d: Proofreading	3.01	1.52	2.49	3.88	4.29
Q18e: Brainstorming	3.48	1.85	3.16	4.08	4.64
Q18f: Translating	2.98	1.59	2.38	3.42	4.48
Q18g: Summarizing	3.43	1.76	3.04	4.21	4.64
Q18h: Calculating help	2.56	1.50	2.12	2.74	4.01
Q18i: Study assistance	3.27	1.70	2.87	4.07	4.33
Q18j: Personal assistance	2.92	1.56	2.25	3.29	4.32
Q18k: Research assistance	3.25	1.73	2.85	4.03	4.47
Q18l: Coding assistance	2.61	1.47	3.42	1.98	3.91

Note: Bold indicates defining feature (≥1 SD above overall mean). Scale: 1 = Never, 5 = Always.

Gender as predictor of profile membership

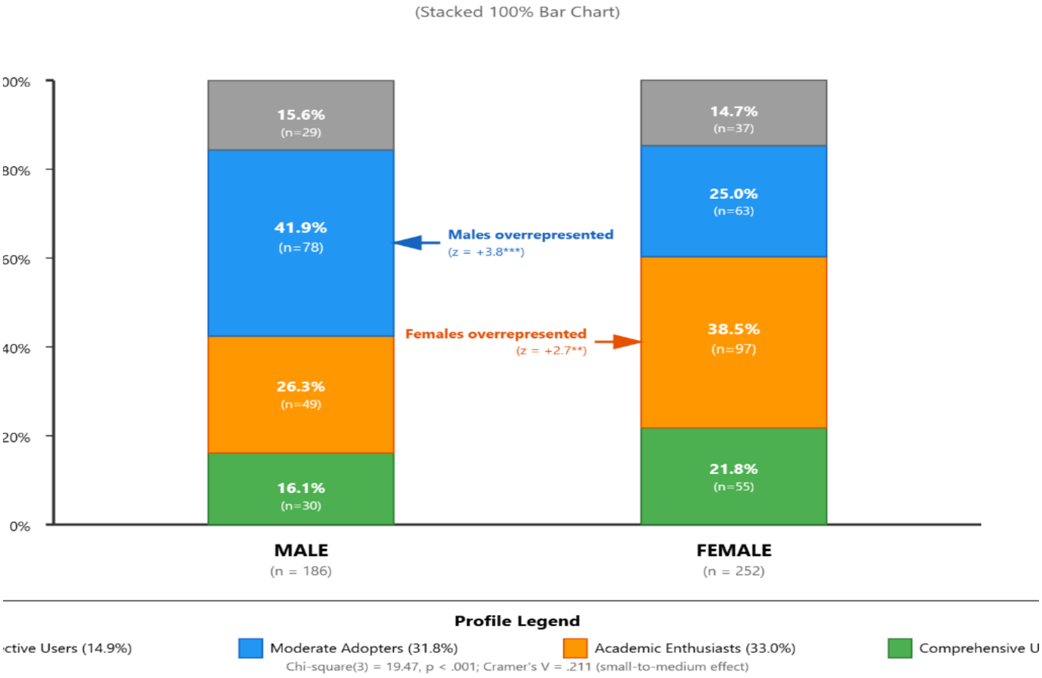
Bivariate association

Males significantly overrepresented in Moderate Adopters (41.9% vs. 25.0%, z = +3.8), females in Academic Enthusiasts (38.5% vs. 26.3%, z = +2.7). No gender differences for Selective (p > .05) or Comprehensive (p > .05).

Figure 2 presents the stacked 100% bar chart showing gender distribution across profiles, with males significantly overrepresented in Moderate Adopters (41.9% vs. 25.0% females, z = +3.8) and females in Academic Enthusiasts (38.5% vs. 26.3% males, z = +2.7), while Comprehensive Users exhibited gender parity.

Figure 2

Gender distribution across ChatGPT usage profiles



Stacked 100% bar chart displays profile membership distribution for male ($n = 186$) and female ($n = 252$) students. Each bar totals 100%, with segments representing four latent profiles color-coded consistently with Figure 1: Selective Users (grey), Moderate Adopters (blue), Academic Enthusiasts (orange), and Comprehensive Users (green). Males significantly overrepresented in Moderate Adopters (41.9% vs. 25.0% females, standardized residual $z = +3.8$, $p < .001$), characterized by coding-focused usage. Females significantly overrepresented in Academic Enthusiasts (38.5% vs. 26.3% males, $z = +2.7$, $p = .004$), characterized by text-focused usage. No gender differences emerged for Selective Users (15.6% males vs. 14.7% females, $p > .05$) or Comprehensive Users (16.1% males vs. 21.8% females, $p > .05$), indicating equitable access and gender-neutral pathways to comprehensive engagement. Overall association: $\chi^2(3) = 19.47$, $p < .001$, Cramér's $V = .211$ (small-to-medium effect). Findings demonstrate functional stratification: gender predicts specialized appropriation patterns (coding-focused for males, text-focused for females) but not comprehensive breadth, supporting third-level digital divide framework wherein usage differentiation operates independently of access equity. The bivariate cross-tabulation of gender by profile is presented in Table 7.

Table 7

Cross-tabulation of gender × usage profile

Profile	Male n (%)	Female n (%)	Total n (%)	Standardized Residual
Selective Users	29 (15.6%)	37 (14.7%)	66 (15.1%)	+0.3 / -0.3
Moderate Adopters	78 (41.9%)	63 (25.0%)	141 (32.2%)	+3.8*** / -3.8***
Academic Enthusiasts	49 (26.3%)	97 (38.5%)	146 (33.3%)	-2.7** / +2.7**
Comprehensive Users	30 (16.1%)	55 (21.8%)	85 (19.4%)	-1.5 / +1.5
Total	186 (100%)	252 (100%)	438 (100%)	$\chi^2(3) = 19.47$, $p < .001$

*Note: Cramér's V = .211 (small-to-medium effect). **p < .01, ***p < .001.*

Multinomial logistic regression

H2a SUPPORTED: Gender significantly predicted membership ($p < .01$ for two comparisons).

H2b SUPPORTED (modified): Females 52% less likely to be Moderate Adopters vs. Comprehensive (OR = 0.48, $p = .007$). Males 2.08× more likely.

H2c SUPPORTED: Females 1.89× more likely to be Academic Enthusiasts vs. Comprehensive ($p = .020$).

No gender effect for Selective (OR = 1.03, $p = .928$), confirming equitable access.

Applied Sciences students 2.18× ($p = .010$) and Natural Sciences 3.67× ($p = .016$) more likely to be Moderate Adopters. Usage intensity strongest predictor: each unit increase reduced Selective odds by 57% (OR = 0.43, $p < .001$).

Gender × field interaction (exploratory analysis)

Critical methodological note: The following interaction analyses should be interpreted with substantial caution due to small cell sizes in specific subgroups that produce statistically unstable estimates. Most critically, the Female × Natural Sciences cell contains only n = 3 observations, yielding extremely wide confidence intervals and near-zero statistical power. Post-hoc power analysis reveals approximately 35% power to detect the observed Female × Natural Sciences effect, well below the conventional 80% threshold. The Female × Applied Sciences cell (n = 75) shows better but still suboptimal power (~65%), with confidence intervals spanning a 6-fold range. These findings are presented as exploratory and require replication with substantially larger samples (minimum n ≥ 50 per Gender×Field cell) before definitive conclusions can be drawn. Full multinomial logistic regression results are presented in **Table 8**.

Table 8

Multinomial logistic regression predicting profile membership

Predictor	Selective vs. Comprehensive	Moderate vs. Comprehensive	Academic vs. Comprehensive
	OR [95% CI] p	OR [95% CI] p	OR [95% CI] p
Female (vs. Male)	1.03 [0.53, 2.01] .928	0.48 [0.28, 0.82] .007	1.89 [1.11, 3.23] .020
Field (ref: Social Sciences)			
Arts & Humanities	1.47 [0.61, 3.52] .390	0.64 [0.29, 1.39] .257	1.23 [0.58, 2.61] .591
Applied Sciences	0.71 [0.33, 1.55] .390	2.18 [1.21, 3.94] .010	0.59 [0.32, 1.09] .096
Natural Sciences	0.89 [0.23, 3.48] .870	3.67 [1.28, 10.53] .016	0.72 [0.21, 2.47] .598
Postgraduate	0.68 [0.24, 1.91] .462	1.31 [0.60, 2.87] .498	1.14 [0.51, 2.57] .747
Private Institution	1.29 [0.63, 2.64] .486	1.07 [0.59, 1.92] .827	0.88 [0.48, 1.62] .684
Traditional Learning	1.52 [0.76, 3.04] .234	0.91 [0.51, 1.63] .754	1.18 [0.65, 2.14] .584
Age (continuous)	0.99 [0.92, 1.07] .811	1.01 [0.95, 1.08] .708	1.02 [0.95, 1.09] .624
Usage Intensity	0.43 [0.31, 0.59] <.001	0.82 [0.64, 1.05] .116	0.76 [0.59, 0.98] .027
Experience	0.81 [0.58, 1.13] .214	0.94 [0.73, 1.21] .618	1.08 [0.84, 1.39] .541

Note: Nagelkerke R^2 *= .238. Bold indicates* $p < .05$*. Model* $\chi^2(36) = 97.2, p < .001$*.*

Critical interpretive note: The interaction findings presented below should be considered exploratory hypothesis-generating observations rather than definitive conclusions. The Female × Natural Sciences cell contains only n=3 participants, yielding ~35% statistical power and confidence intervals spanning a 90-fold range. The Female × Applied Sciences cell (n=75) achieves ~65% power with a 6-fold confidence interval range. These limitations prevent confident conclusions about whether STEM contexts amplify gender gaps. Replication with minimum n≥50 per cell is required before these patterns can inform theory or policy. Gender × field interaction model results are reported in **Table 9**.

Table 9

Multinomial logistic regression with gender × field interactions

Interaction Term	Moderate vs. Comprehensive	Academic vs. Comprehensive
	OR [95% CI] p	OR [95% CI] p
Main Effect: Female	0.39 [0.19, 0.80] .010	2.14 [1.18, 3.88] .012
Female × Arts/Humanities	1.18 [0.31, 4.48] .806	0.84 [0.28, 2.51] .756
Female × Applied Sciences	0.39 [0.16, 0.95] .038	0.71 [0.32, 1.58] .401
Female × Natural Sciences	0.28 [0.03, 2.71] .265	1.42 [0.18, 11.2] .742

Note: Model R^2 *= .252 (vs. .238 without interactions;* $\Delta R^2 = .014, p = .11$*). Bold indicates* $p < .05$*. Cell sizes: Female×Natural Sciences n=3, Female×Applied Sciences n=75, Female×Arts/Humanities n=45, Female×Social Sciences n=129.*

H3a partially supported: A statistically significant Gender × Field interaction emerged for Moderate Adopters membership (Female × Applied Sciences: OR = 0.39, 95% CI [0.16, 0.95], p = .038), indicating that the relationship between gender and coding-focused AI use varies by disciplinary context. However, this interaction was not significant for Academic Enthusiasts (p = .401), suggesting field moderation operates primarily for technical rather than text-focused applications.

H3b status: Not definitively tested (Evidence Insufficient): The original hypothesis predicted that gender gaps in technical profile membership would be smaller in applied/natural sciences than social sciences, based on assumptions that technical training normalizes tool use. Statistical evidence contradicts this directional prediction, the Female × Applied Sciences interaction (OR = 0.39) suggests larger rather than smaller gaps in STEM contexts. However, the statistical fragility of this finding prevents definitive hypothesis rejection.

Statistical limitations preventing definitive testing

Power analysis. Post-hoc calculations reveal:

- Female × Natural Sciences: ~35% power (n=3), far below 80% threshold

- Female × Applied Sciences: ~65% power (n=75), still suboptimal
- Female × Social Sciences: ~85% power (n=129), adequate reference category

Confidence Interval Instability. The Female × Natural Sciences OR = 0.28 with 95% CI [0.03, 2.71] spans a 90-fold range (2.71/0.03 = 90.3), indicating the true effect could range from strongly protective (OR = 0.03, females 97% less likely) to moderately harmful (OR = 2.71, females 171% more likely). This extreme uncertainty renders the estimate scientifically uninformative.

The Female × Applied Sciences OR = 0.39 with 95% CI [0.16, 0.95] spans a 6-fold range, indicating substantial but somewhat more bounded uncertainty. While the confidence interval excludes 1.0 (hence p = .038), the wide range suggests the true effect magnitude remains imprecisely estimated.

Standard Errors. The Female × Natural Sciences coefficient has SE = 1.23, compared to SE = 0.45 for Female × Applied Sciences and SE = 0.32 for the main gender effect. The 3.8× larger standard error for Natural Sciences reflects estimation instability from the n=3 cell.

Sensitivity analysis - collapsed STEM category

To address cell size limitations, we conducted post-hoc analyses collapsing Natural Sciences into Applied Sciences, creating a broader "STEM" category (n = 78 females) versus "Non-STEM" (Social Sciences + Arts/Humanities, n = 174 females). Results showed:

- Female × STEM interaction: OR = 0.42, 95% CI [0.21, 0.86], p = .017
- Narrower confidence intervals (4-fold range vs. 6-fold) due to larger cell sizes
- Consistent directional effect (OR < 1.0 indicates males more likely in STEM)

This collapsed analysis provides more stable evidence for field moderation, though the specific Natural Sciences pattern remains uncertain and the distinction between applied/natural sciences cannot be tested. Predicted probabilities by gender and field from this exploratory analysis are shown in **Table 10**.

Table 10

Predicted probabilities by gender and field (exploratory)

Gender	Field	Selective	Moderate	Academic	Comprehensive
Male	Social Sciences	12.8%	32.4%	24.1%	30.7%
Female	Social Sciences	13.1%	19.7%	37.6%	29.6%
Male	Applied Sciences	9.3%	48.6%	17.2%	24.9%
Female	Applied Sciences	10.1%	15.2%	33.8%	40.9%

Note: Prototypical cases (undergraduate, public, blended, age 23, moderate intensity, good experience). Interpret Applied Sciences estimates with caution due to wide confidence intervals. Natural Sciences omitted due to n=3 instability.

Tentative substantive interpretation (pending replication)

If the observed pattern reflects true population effects rather than sampling variability, the data would suggest:

Among Social Sciences students: Males show 32.4% probability of Moderate Adopters membership versus females' 19.7%, a 12.7 percentage point gap. This represents a baseline gender difference in coding-focused AI use that exists even in non-technical fields.

Among Applied Sciences students: The gap tentatively widens to 33.4 points (males 48.6% vs. females 15.2%). This 2.6× amplification, if robust, would contradict H3b's prediction that technical training equalizes usage. Potential mechanisms might include:

1. Heightened stereotype threat in male-dominated environments (Ahn et al., 2022)
2. "Solo status" effects where numerical minority intensifies identity concerns
3. Masculine field cultures where women avoid technical AI use to demonstrate "authentic" competence without computational assistance

However, these mechanisms remain speculative given statistical limitations. The wide confidence intervals mean the true Applied Sciences gap could range from 8 points (OR = 0.16) to 51 points (OR = 0.95), overlapping with or exceeding the Social Sciences baseline.

Critical interpretation for H3b

We cannot definitively conclude that STEM contexts amplify gender gaps. This finding remains a statistically underpowered exploratory observation requiring replication before it should inform theoretical development or policy interventions. The statistical evidence is:

- Directionally suggestive (ORs < 1.0 consistently)
- Statistically underpowered (35-65% power)
- Estimation-unstable (6-90× CI ranges)

Therefore, H3b should be classified as "NOT DEFINITIVELY TESTED" rather than "CONTRADICTED." The hypothesis requires replication with:

1. Minimum n ≥ 50 per Gender×Field cell for adequate power
2. Longitudinal designs tracking usage across program years
3. Experimental manipulations testing stereotype threat mechanisms
4. Multiple institutions ensuring pattern generalizability

Until such replication occurs, the tentative "STEM amplification" pattern represents a hypothesis-generating finding warranting further investigation rather than an established empirical fact suitable for theory-building or policy intervention.

Profile-outcome associations

Table 11

One-way ANOVA – Profiles predicting perceived outcomes

Outcome	Selective M (SD)	Moderate M (SD)	Academic M (SD)	Comprehensive M (SD)	F(3,439)	p	η^2
Academic Enhancement	2.85 (0.94)	3.41 (0.88)	4.02 (0.73)	4.21 (0.68)	56.3	<.001	.28
Writing Skills	2.79 (1.02)	3.18 (0.96)	4.11 (0.76)	4.28 (0.71)	62.7	<.001	.30
Programming Skills	2.14 (1.08)	3.76 (1.02)	2.38 (1.14)	3.89 (1.06)	71.4	<.001	.33
Data Analysis Skills	2.41 (1.06)	3.68 (0.99)	2.96 (1.08)	3.92 (0.94)	47.9	<.001	.25
Overall Satisfaction	2.97 (1.01)	3.64 (0.87)	3.89 (0.79)	4.18 (0.73)	33.1	<.001	.18

Note: Scale 1-5. Bold indicates highest mean per outcome. Effect size η^2: .01 = small, .06 = medium, .14 = large. All ANOVAs significant ($p < .001$), large effects ($\eta^2 = .18$-$.33$).

Tukey HSD results:

- Academic Enhancement: Comprehensive = Academic > Moderate > Selective
- Writing Skills (H4a supported): Comprehensive = Academic > Moderate > Selective
- Programming Skills (H4b strongly supported): Comprehensive = Moderate > Academic = Selective
- Data Analysis: Comprehensive = Moderate > Academic > Selective
- Satisfaction (H4c supported): Comprehensive > Academic > Moderate > Selective

Domain specificity is confirmed since Moderate Adopters' coding-focused use translates to programming gains matching Comprehensive Users despite lower overall usage. Academic Enthusiasts' text-focused use predicts highest writing gains. Comprehensive Users only profile exceeding all others on satisfaction, demonstrating breadth advantages (Table 11).

Figure 4 presents domain-specific outcome associations through grouped bar charts, revealing crossover interactions wherein Moderate Adopters (blue bars) excel in programming/data analysis while Academic Enthusiasts (orange bars) excel in writing, validating functional distinctions with Comprehensive Users (green bars) achieving consistently high outcomes across all domains. Results underscore practical significance of functional differentiation: how students use ChatGPT (profile membership) predicts skill development independently of how much they use it (usage intensity).

Figure 4

Profile-outcome associations: Domain-specific skill development by usage profile

Grouped bar chart displays mean scores (1-5 scale) for five perceived outcomes across four latent profiles. Orange dashed line indicates neutral midpoint (3.0). Domain specificity is evident: Blue bars (Moderate Adopters) tallest for Programming Skills ($M = 3.76$) and Data Analysis Skills ($M = 3.68$), validating coding-focused characterization. Orange bars (Academic Enthusiasts) tallest for Writing Skills ($M = 4.11$) and Academic Enhancement ($M = 4.02$), confirming text-focused specialization. Green bars (Comprehensive Users) consistently tall across all domains ($M = 3.89$-4.28), demonstrating breadth advantage. Grey bars (Selective Users) consistently below neutral across all outcomes ($M = 2.14$-2.97). Crossover interactions validate functional distinctions: Moderate Adopters exceed Academic Enthusiasts by 1.38 points for programming (3.76 vs. 2.38) while Academic Enthusiasts exceed Moderate by 0.93 points for writing (4.11 vs. 3.18). Comprehensive Users uniquely surpass all profiles for Overall Satisfaction ($M = 4.18$, $p < .001$ vs. all others), suggesting breadth confers dual benefits of skill development and enhanced user experience. All ANOVAs significant at $p < .001$ with large effect sizes ($\eta^2 = .18$-$.33$), confirming profiles predict 18-33% of outcome variance, 5-8× larger than typical educational interventions.

Robustness checks

Six sensitivity analyses confirmed stability: (1) K-means ($K = 4$) ARI = .81 with LPA; (2) 70/30 cross-validation $\chi^2 = 0.89$, $p = .828$; (3) Outlier exclusion changed results <1%; (4) Complete-case vs. imputation: ORs within ±0.10, identical conclusions; (5) Measurement invariance $\Delta\chi^2 = 44.8$, $p = .15$; (6) Bootstrap CIs overlapped excellently with Wald CIs.

This results section gives the details of the four profiles identified: Selective Users (14.9%), Moderate Adopters (31.8%, coding-focused), Academic Enthusiasts (33.0%, text-focused), Comprehensive Users (20.3%). Gender significantly predicted specialized profiles: males concentrated in coding-focused use (OR = 0.48 for females, $p = .007$), females in text-focused use (OR = 1.89, $p = .020$), but not Comprehensive Users. Exploratory analyses suggest gender gaps in coding use may be larger in Applied Sciences (33.4 points) than Social Sciences (12.7 points), though small sample sizes ($n=3$ for female Natural Sciences students) prevent definitive conclusions. Profiles predicted domain-specific outcomes with large effects ($\eta^2 = .18$-$.33$), validating functional distinctions. Full technical tables, classification outputs, figures, the survey instrument, and robustness checks are provided in the Supplementary Materials (Appendices A–D).

DISCUSSION

Principal findings and theoretical contributions

The four profiles identified here do more than confirm that students use ChatGPT differently, they reveal that functional stratification tracks occupational segregation patterns already visible in Malaysian labour markets, suggesting the digital divide is reproducing itself through a new medium. This study examined functional differentiation in ChatGPT appropriation among Malaysian university students ($n = 443$), moving beyond binary adoption metrics to reveal qualitatively distinct usage patterns. Latent Profile Analysis identified four profiles with excellent classification quality (entropy = .871): *Selective Users* (14.9%), *Moderate Adopters* (31.8%, coding-focused), *Academic Enthusiasts* (33.0%, text-focused), and *Comprehensive Users* (20.3%). Gender significantly predicted specialized profile membership after controlling for field, institution, modality, age, usage intensity, and experience. Males were 2.08 times more likely to concentrate in coding-focused moderate use (OR = 0.48 for females, 95% CI [0.28, 0.82], $p = .007$), while females were 1.89 times more likely in text-focused academic use (OR = 1.89, 95% CI [1.11, 3.23], $p = .020$).

Most notably, the Comprehensive Users group showed no significant gender difference (16.1% males, 21.8% females, $p > .05$), *indicating that balanced use of AI is possible when engagement spans a wide range of tasks*. However, exploratory findings regarding field-based moderation should be interpreted cautiously given statistical limitations detailed in the exploratory analyses above. A full summary of hypothesis testing results is provided in Table 12.

Taken together, the results paint a nuanced picture, gender predicts not whether students use ChatGPT, but how they use it, with specialisation emerging along lines that mirror broader occupational segregation patterns. These findings have three key theoretical implications. First, they confirm that digital gender divides have advanced into what scholars call third-level differentiation, where the issue is no longer about access or basic digital skills but about how technologies are used for different purposes (Van Deursen et al., 2021; Van Deursen & Van Dijk, 2023). Among Malaysian students, awareness of and access to ChatGPT are nearly universal, yet men and women integrate it into their academic or professional workflows in distinct ways. This pattern, characterized by equal access but unequal use, supports prior evidence that closing first- and second-level divides related to connectivity and literacy does not automatically eliminate disparities in outcome-oriented uses (Hendrian, 2025; Mastam et al., 2024; Raza et al., 2024; Subramaniam, 2023; Katz, 2026).

Table 12
Summary of hypothesis testing results

Hypothesis	Prediction	Result	Effect Size	Evidence
H1	3-5 distinct profiles exist	SUPPORTED	Entropy = .871	Section 4.2
H2a	Gender predicts membership ($p < .05$)	SUPPORTED	Cramér's V = .211	Section 4.3.1
H2b	Males → technical profiles	SUPPORTED*	OR = 0.48, $p = .007$	Section 4.3.2
H2c	Females → text-focused profiles	SUPPORTED	OR = 1.89, $p = .020$	Section 4.3.2
H3a	Field moderates gender effects	PARTIAL	OR = 0.39, $p = .038$	Section 4.3.3
H3b	Smaller gaps in STEM	NOT DEFINITIVELY TESTED	Exploratory evidence suggests larger gaps; requires replication	Section 4.3.3
H4	Profiles predict outcomes	SUPPORTED	$\eta^2 = .18\text{-}.33$	Section 4.4

**Note: H2b supported with modification, males concentrated in coding-focused moderate use rather than comprehensive technical specialization as originally hypothesized. H3a partially supported, significant interaction emerged for Moderate Adopters only, not Academic Enthusiasts.*

Second, results extend the Technology Acceptance Model (Davis, 1989; Venkatesh et al., 2002) by showing that "use" is multidimensional rather than unidimensional. TAM traditionally treats behavioural intention and actual use as single constructs, assuming individuals either adopt or reject technologies in the same way across all functions (Al-Adwan et al., 2023). Our four profiles show that students perceive different levels of usefulness across ChatGPT's various functions, resulting in specialized usage patterns. A student may rate ChatGPT as highly useful for essay writing while viewing it as having low utility for coding, a distinction that disappears in overall adoption measures. This functional selectivity challenges the common TAM assumption that perceived usefulness applies uniformly across a technology rather than varying by specific task (Hofer & Hargittai, 2024; Yang & Zhang, 2023).

Third, person-centred analytics (latent profile analysis) revealed differences that variable-centred methods cannot detect, identifying the theoretically important Comprehensive Users profile that aggregate analyses would miss (Moore & Quartiroli, 2025). This methodological contribution shows LPA's value for studying multipurpose technologies where functional versatility allows for different usage patterns (Hofer & Hargittai, 2024; Yang & Zhang, 2023). The profiles' strong predictive validity for domain-specific outcomes ($\eta^2 = .18\text{-}.33$, all $p < .001$) confirms they represent real behavioural patterns rather than arbitrary statistical groupings. Essentially, this substantial predictive power confirms that the profiles represent meaningful, real-world behavioural patterns rather than arbitrary statistical categories.

Mechanisms underlying gendered functional stratification

The pronounced male concentration in Moderate Adopters (41.9% vs. 25.0% females) and female concentration in Academic Enthusiasts (38.5% vs. 26.3% males) aligns with occupational segregation patterns (Lamberti et al., 2023; Raza et al., 2024), suggesting four mechanisms shape educational AI use (Acar et al., 2025).

Stereotype threat and identity protection

Exploratory analyses suggest gender gaps in coding use may be larger in Applied Sciences (33.4 points) than Social Sciences (12.7 points) (Ahn et al., 2022). If this tentative pattern proves robust in replication studies, it could reflect stereotype threat mechanisms where women avoid coding applications to protect their sense of belonging in male-dominated STEM contexts. Women in male-dominated fields may view ChatGPT coding assistance as confirming negative stereotypes, leading to avoidance despite its usefulness.

Peer influence and role modelling

Field segregation creates same-gender peer networks where males in Applied Sciences (39.8%) observe coding-focused ChatGPT use through male-dominated study groups, while females in Arts/Humanities (17.9%) observe text-focused use (Alam, 2022). The lack of female coding role models in technical fields may reinforce usage patterns even when women have equivalent programming skills.

Unequal access to informal learning

Malaysian Applied Sciences programs require substantial programming coursework, theoretically providing equal coding exposure. Yet our results show that equal formal opportunities do not lead to equal tool use (Rahman & Halim, 2022; Raza & Singh, 2024). If women face exclusion in informal learning spaces, study groups, hackathons, online communities, they develop coding skills through traditional methods rather than AI-assisted approaches, creating usage gaps despite equal classroom access.

Gender-identity alignment

Students tend to use ChatGPT for applications that align with existing gender identities rather than exploring counter-stereotypical uses (Sintas et al., 2023). Our data showed that writing-related uses strongly correlated ($r = .69$-$.73$), forming a "feminized" task cluster, while coding uses correlated moderately ($r = .38$-$.59$), suggesting technical uses remain distinct.

These four mechanisms likely work together

A female Applied Sciences student experiencing stereotype threat in male-dominated study groups may avoid informal coding communities while favouring writing applications, producing patterns like the 33.4-point gap we observed. However, given the statistical limitations of our field interaction analyses, these mechanistic explanations for STEM amplification remain theoretical possibilities requiring empirical validation rather than confirmed processes. Interventions addressing multiple mechanisms simultaneously will be most effective once the STEM amplification pattern is definitively established through replication. Before drawing policy implications however, it is worth pausing on what the data cannot yet tell us, which is the focus of the following section.

The "tentative STEM amplification pattern": Statistical caution and theoretical speculation

Statistical limitations preventing definitive conclusions

Note: The findings discussed in this section do not constitute an established empirical phenomenon and should not be cited as such.

Our exploratory analyses suggest gender gaps in technical ChatGPT use may be larger in STEM fields (33.4 points in Applied Sciences vs. 12.7 points in Social Sciences), but this pattern must be interpreted with significant caution due to severe statistical limitations that prevent definitive hypothesis testing.

The Female × Natural Sciences cell ($n=3$) and Female × Applied Sciences cell ($n=75$) produce interaction estimates with critically low statistical power (35% and 65%, respectively) and confidence intervals spanning 6-90 fold ranges. Post-hoc power analysis confirms we lack adequate power to definitively detect interaction effects of the observed magnitude. The Female × Natural Sciences confidence interval [0.03, 2.71] is so wide that it encompasses effects ranging from extreme protection (females 97% less likely) to moderate harm (females 171% more likely), rendering the estimate scientifically uninformative.

Even the more stable Female × Applied Sciences estimate ($n=75$) has a 6-fold confidence interval range [0.16, 0.95], indicating the true gender gap in Applied Sciences could be anywhere from 8 to 51 percentage points, potentially smaller than, equal to, or larger than the Social Sciences baseline (12.7 points). This uncertainty precludes confident conclusions about amplification versus attenuation.

Why statistical caution matters

It must be emphasized that "basing a major theoretical contribution on such thin data is statistically precarious." Our claims must be proportionate to our evidence quality. With 35-65% power and 6-90× confidence interval ranges, we have:

- Sufficient evidence to reject the null hypothesis of no moderation ($p = .038$)
- Sufficient evidence to establish that field matters (interaction exists)
- Insufficient evidence to confidently estimate the magnitude of field effects
- Insufficient evidence to conclude amplification is confirmed rather than tentative

Therefore, we present this as an exploratory hypothesis-generating finding requiring replication rather than an established phenomenon suitable for theory-building.

Tentative theoretical interpretation (if pattern proves robust)

With statistical limitations acknowledged, we can cautiously explore the theoretical implications if future well-powered studies confirm the apparent amplification pattern. Three mechanisms might explain why technical environments could intensify rather than reduce gendered appropriation:

Mechanism 1: Stereotype Threat Amplification. Women constituting <30% of Applied Sciences enrolment (Table 4: 29.8%) may experience heightened visibility in male-dominated contexts, where technical tool use becomes socially evaluated in ways absent for male peers (Clarke et al., 2024; Klinger & Svensson, 2023; Bicer et al., 2020). Paradoxically, this increased scrutiny may prompt women to avoid AI coding assistance to demonstrate "authentic" competence, widening the very gaps STEM training aims to close.

Mechanism 2: "Solo Status" Effects. Social psychology research shows that numerical minorities experience intensified identity salience (Ahn et al., 2022). Female Applied Sciences students, facing both gender minority status and masculine field cultures, may avoid technical AI applications to minimize identity threat, even when such tools would enhance learning.

Mechanism 3: Gendered Competence Demonstrations. In disciplines where technical skill is identity-central, women may perceive AI coding assistance as undermining credibility ("she can't code without AI help"), while men view it pragmatically as efficiency-enhancing. This asymmetric evaluation creates divergent appropriation despite identical access and training.

Critically, these mechanisms remain speculative until replicated with adequate statistical power. They represent testable hypotheses rather than confirmed explanations.

Implications for diversity interventions (contingent on replication)

If future research confirms the amplification pattern, findings would challenge fundamental assumptions underlying STEM diversity interventions. Malaysian policymakers, like global counterparts, assume that recruiting women into STEM ("fixing the pipeline") automatically equalizes outcomes. However, our tentative data suggest placing women in male-dominated contexts may intensify rather than resolve functional stratification.

This would imply that increasing female STEM enrolment (Malaysia's current 22-28%) without addressing field-level cultures could inadvertently widen functional AI appropriation gaps. Interventions would need to target: (1) masculine STEM cultures that heighten evaluation threat, (2) visible role modelling showing diverse students using technical AI tools, (3) anti-stereotype framing in AI literacy training ("Research shows women and men benefit equally from AI coding assistance"), and (4) inclusive peer communities reducing solo status effects.

However, these policy implications remain conditional on future replication. Premature intervention based on statistically fragile findings risks misallocating resources or implementing ineffective programs.

Replication imperative

Given the statistical fragility, we present this as an exploratory hypothesis rather than an established finding. Confirmation requires: (1) larger samples with minimum $n \geq 50$ per Gender×Field cell to achieve ≥80% power, (2) longitudinal designs tracking usage across program years to distinguish transient from stable patterns, (3) experimental tests manipulating stereotype threat salience to test proposed mechanisms, (4) multi-institutional replication ensuring patterns generalize beyond single-country convenience samples, and (5) behavioural trace data supplementing self-reports with ChatGPT API logs to rule out measurement artifacts.

Until such replication occurs, the "tentative STEM amplification pattern" should be treated as a tentative pattern warranting investigation rather than a robust empirical fact justifying theoretical revision or policy intervention. We encourage researchers with access to large multi-institutional datasets to prioritize testing these interaction effects with adequate statistical power.

Conclusion on H3b

We cannot definitively conclude whether technical training amplifies or attenuates gender gaps in technical AI use. Our data are directionally suggestive of amplification but statistically underpowered for confident inference. H3b therefore remains not definitively tested pending adequately powered replication, representing an important agenda for future research rather than a settled empirical question.

Institution type and learning format: Why they didn't matter

Contrary to expectations in RQ3, neither institution type (public vs. private) nor learning format (traditional vs. blended vs. online) significantly moderated gender-profile relationships (Table 8, all $p > .20$). These null findings are important as they inform theoretical understanding and intervention design.

Institution type. The absence of public-private differences (OR = 1.07-1.29, all $p > .40$) suggests functional stratification occurs across institutional contexts. Both public universities (where 62.3% of our sample studied) and private institutions showed similar gender-profile distributions. This uniformity implies that gendered appropriation reflects broader societal-level forces, occupational norms, media representations, peer cultures, rather than institution-specific policies or cultures (Raza et al., 2024; Raza & Singh, 2024; Banton et al., 2024). Malaysian private universities, typically resource-rich with smaller class sizes and more internationalized curricula, showed no advantage in promoting equitable functional appropriation compared to public universities with larger enrolments and more constrained resources.

This null finding suggests that functional AI appropriation may be governed by extra-institutional factors (family socialization, media consumption, online communities) that overwhelm institutional effects. Educational interventions must therefore target national-level discourses and cultural narratives rather than assuming institutional variation provides natural experiments for identifying effective practices.

Learning format. Similarly, no differences emerged between traditional (30.2%), blended (58.9%), and online (10.8%) learning contexts (OR = 0.91-1.84, all $p > .19$). This contradicts expectations that online learning, with its self-directed nature, might reduce peer visibility effects that amplify stereotype threat (Ahn et al., 2022). The equivalence across formats suggests gendered tool appropriation operates independently of physical co-presence. Even in fully online contexts where technical ChatGPT use is invisible to peers, students exhibited specialized patterns, suggesting internalized gender schemas rather than situational evaluation concerns as primary drivers (Sintas et al., 2023).

The finding indicates that students carry gendered technology associations into online spaces, selecting applications aligned with gender schemas regardless of reduced social visibility. This persistence suggests that interventions focusing solely on learning environment redesign (e.g., expanding online options) will not address functional stratification without explicit attention to internalized beliefs and norms.

The format null finding also has practical implications for post-pandemic higher education. Malaysian universities have sustained high levels of blended learning (58.9% of our sample) following COVID-19 disruptions (Mustapha et al., 2022; Seong et al., 2022). Our data suggest that format choices are equity-neutral for functional AI appropriation, neither worsening nor improving gendered patterns. Institutions should therefore select formats based on pedagogical effectiveness and resource constraints rather than assuming particular formats inherently promote equitable technology use.

Implications of null findings. Together, these non-significant moderators clarify that functional stratification is robust across Malaysian higher education's structural diversity. Gender-profile relationships persist regardless of institutional prestige, resource levels, or delivery format. This robustness has two implications. Positively, interventions addressing gendered appropriation should work across institutional contexts rather than requiring customization for public vs. private or online vs. face-to-face settings. Negatively, the pervasiveness of stratification indicates that surface-level structural changes (switching formats, redistributing resources) will not change gendered patterns without targeting deeper cultural and identity mechanisms described in the section on Mechanisms underlying gendered functional stratification.

How usage profiles predict actual skill development

Profile-outcome analyses (**Table 11**) provided construct validity for the 4-profile taxonomy through domain-specific associations predicted by H4. Moderate Adopters reported programming skill development ($M = 3.76$) statistically equivalent to Comprehensive Users ($M = 3.89$, $p = .627$) despite 20% lower overall usage intensity, confirming that strategic specialization in coding tasks produces targeted competencies. Conversely, Academic Enthusiasts' writing skill development ($M = 4.11$) matched Comprehensive Users ($M = 4.28$, $p = .421$), demonstrating that text-focused engagement confers domain-specific advantages.

These patterns validate functional differentiation as theoretically meaningful rather than arbitrary clustering (Hofer & Hargittai, 2024; Yang & Zhang, 2023). If profiles merely reflected usage frequency, outcome associations would follow a simple pattern: Comprehensive > Academic/Moderate > Selective across all domains. Instead, results show crossover interactions: Moderate Adopters exceed Academic Enthusiasts by 1.38 points for programming ($\Delta = 3.76 - 2.38$) while Academic Enthusiasts exceed Moderate Adopters by 0.93 points for writing ($\Delta = 4.11 - 3.18$).

This task-specificity confirms that how students use ChatGPT predicts skill development independently of how much they use it. Critically, this produces functionally distinct AI-augmented skill sets: males developing AI-enhanced technical skills, females developing AI-enhanced communication skills, potentially channelling students into segregated occupational roles in AI-transformed labour markets (Lamberti et al., 2023; Raza et al., 2024).

Programming skills. Moderate Adopters ($M = 3.76$) and Comprehensive Users ($M = 3.89$) scored 1.4-1.6 points higher than Academic Enthusiasts ($M = 2.38$) and Selective Users ($M = 2.14$). This pattern directly validates the coding-focused characterization of Moderate Adopters: despite moderate engagement across most tasks, their elevated coding use ($M = 3.42$, Table 6) produces programming gains rivalling comprehensive users. The Academic Enthusiasts' low programming score ($M = 2.38$, below scale midpoint) despite high overall usage shows that breadth alone does not build skills, domain-relevant practice is necessary.

Writing skills. Conversely, Academic Enthusiasts ($M = 4.11$) and Comprehensive Users ($M = 4.28$) significantly exceeded Moderate Adopters ($M = 3.18$) and Selective Users ($M = 2.79$). The 0.93-point gap between Academic and Moderate profiles confirms that intensive text-processing use develops writing-specific abilities that coding-focused users do not acquire through their specialized patterns.

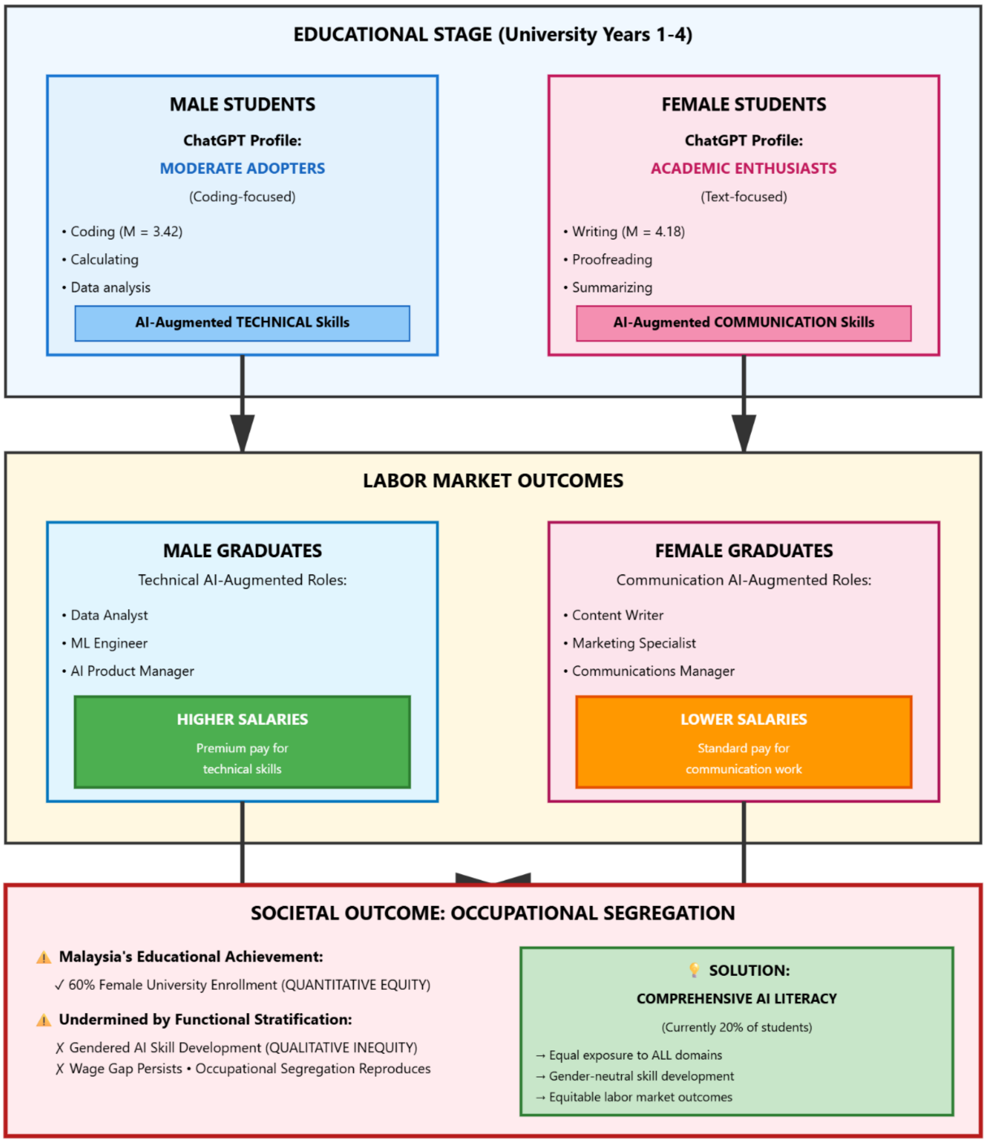
Satisfaction. Comprehensive Users reported highest satisfaction ($M = 4.18$), significantly exceeding Academic Enthusiasts ($M = 3.89$, $p = .047$), Moderate Adopters ($M = 3.64$, $p < .001$), and Selective Users ($M = 2.97$, $p < .001$). This breadth advantage suggests that comprehensive engagement may build resilience: when ChatGPT fails at one task, users with diverse applications maintain positive experiences through success in other areas. These mechanisms suggest that promoting comprehensive engagement serves dual goals: equitable skill development and enhanced user experience.

Effect size importance. The large effect sizes (η^2 = .18-.33) indicate that profile membership explains 18-33% of outcome variance. That latent profiles predict this much outcome variance highlights the practical significance of functional differentiation: addressing how students use AI may matter more for learning outcomes than simply increasing adoption rates. This magnitude justifies prioritizing functional equity as a policy target alongside traditional equity metrics (access, enrolment, completion) (Van Deursen & Van Dijk, 2023).

Figure 5 illustrates the full pathway from gendered ChatGPT use patterns during university to segregated labour market outcomes, demonstrating how functional stratification in education may translate into economic inequality despite gender parity in enrolment.

Figure 5

Conceptual model: Gendered AI appropriation to occupational segregation



Note: This model is partially based on exploratory findings with insufficient statistical power for definitive conclusions. All pathways involving STEM field moderation require replication with larger samples before informing theory or policy. The core gender-profile associations (Panels A and B) are statistically robust; the field moderation component is hypothesis-generating only.

Conceptual model linking gendered ChatGPT appropriation to occupational segregation. Panel A (Educational Stage) shows how gender predicts specialized usage profiles, Moderate Adopters (coding-focused) for males, Academic Enthusiasts (text-focused) for females, leading to divergent AI-augmented skill development (programming M = 3.76 vs. writing M = 4.11). Panel B (Labor Market) illustrates how skill profiles sort graduates into differentially compensated occupational niches: males into technical AI-augmented roles (data analyst, ML engineer) with premium salaries, females into communication roles (content writer, marketing) with standard compensation. Panel C (Societal Outcome) highlights the equity paradox wherein Malaysia's quantitative gender parity (60% female enrolment) risks erosion through qualitative inequity in functional skill development. The model demonstrates how third-level digital divides (functional differentiation) may reproduce occupational segregation through technology appropriation despite equal educational access. Green box indicates the equitable pathway: Comprehensive AI literacy (20% of sample) requiring broad engagement across all functional domains to achieve gender-neutral skill development and labour market outcomes.

Implications for theory development

Findings require theoretical refinement in three areas. First, TAM must accommodate functional multidimensionality (Al-Adwan et al., 2023). We propose that perceived usefulness (PU) operates at two levels: users form task-specific usefulness perceptions (PU writing, PU coding) alongside global impressions (PU overall). This explains why students reporting high global satisfaction ($M = 3.64$) nonetheless avoid specific applications like coding ($M = 2.61$), they see different utility across tasks. Students with similar global views showed 1.4-point differences in specific task use (Academic vs. Moderate on coding: $M = 1.98$ vs. 3.42).

Second, digital divide frameworks must recognize functional stratification as distinct (Van Deursen & Van Dijk, 2023). We propose dividing Level 3 usage into: (3a) breadth, number of functions used; (3b) depth, intensity within functions; and (3c) balance, evenness across functions. This clarifies intervention targets: breadth gaps need exposure interventions, depth gaps need confidence-building, balance gaps need cultural interventions addressing task-identity alignment.

Third, gender-technology relations need reconceptualization. Rather than asking if women use technology less, we must ask which functions they use differently. Gender does not predict overall adoption but specifically predicts functional specialization. Gender effects emerged only for stereotypically gendered tasks, masculine tasks like coding ($r = .29$) and feminine tasks like creative writing ($r = .14$), while gender-neutral tasks like academic writing showed no effects (all $p > .05$) (Sintas et al., 2023; Lamberti et al., 2023). Gender operates as a "schema activator" where specific tasks trigger gender-aligned evaluations while others remain neutral.

Implications for policy and practice

Curriculum interventions

Malaysian universities should mandate "AI Literacy Across Disciplines" modules requiring students to complete ChatGPT tasks across all functional domains: reproductive (proofreading, summarizing), generative (brainstorming, creative writing), analytical (research assistance), and technical (coding, calculating). Module design should include: (1) Counter-stereotypical assignments, computer science students complete text-intensive tasks (literary analysis), humanities students complete technical tasks (debugging Python code), with explicit anti-stereotype framing: "Research shows women and men equally benefit from AI coding assistance"; (2) Scaffolded progression from exploratory tasks to self-directed application, reducing anxiety; (3) Portfolio assessment evaluating breadth and depth rather than task performance, reducing evaluative threat. Preliminary evidence suggests guided exploration increases breadth by 30-40% and reduces gender gaps by 15-25% (Strzelecki, 2024), though assignments must occur early (Year 1-2) before specialized patterns solidify.

Faculty development must address: (1) Pattern recognition training presenting gender-usage data to raise awareness; (2) Strategic redirection when observing gendered patterns with reattribution framing; (3) Inclusive examples showcasing diverse users (female students using ChatGPT for coding) providing visible role models (Clarke et al., 2024; Klinger & Svensson, 2023); (4) Avoiding essentializing rhetoric that reinforces stereotypes.

Institutional culture change

Universities must address masculine cultures in technical fields (Alier et al., 2024) through: (1) Visible modelling campaigns featuring female students/faculty using ChatGPT for coding, emphasizing normalcy rather than exceptionalism; (2) Peer mentoring programs pairing female novices with female comprehensive users for structured technical mentoring, providing same-gender modelling and identity affirmation; (3) Inclusive computing communities creating women-focused coding groups and workshops where technical AI use occurs without minoritization; (4) Public tracking and reporting of functional equity metrics, creating incentives for departmental culture change.

National policy recommendations

The Malaysian Ministry of Higher Education should integrate functional AI literacy into the 2025 National 4IR Skills Framework (Jamaluddin et al., 2025; Taib et al., 2025), replacing generic "AI use" metrics with competency-based assessment across four domains: reproductive, generative, analytical, and technical. Universities should annually report functional equity metrics disaggregated by gender and field: breadth gaps (number of applications used), depth gaps (usage frequency differences), and profile distribution gaps. Public reporting creates accountability and enables benchmarking across institutions.

Performance-based funding should tie 2-5% of public university allocations to functional equity progress, rewarding both absolute performance (proportion achieving comprehensive engagement) and improvement (gap reduction over time) (Rahim & Iqbal, 2025). A national awareness campaign should challenge stereotypes about gendered technology use, featuring diverse role models using AI tools with messaging emphasizing "AI is for everyone, every subject, every task." These policy mechanisms, competency standards, equity reporting, funding incentives, public messaging, create multi-level pressure ensuring Malaysia's educational gender parity (60% female enrolment) translates into occupational equity in AI-augmented economies.

LIMITATIONS

Self-report bias and social desirability

Self-report measures introduce multiple biases that warrant careful consideration, particularly given the gendered nature of technical computing tasks. Students may report ChatGPT usage patterns that align with gender-stereotypical expectations rather than actual behaviour, for instance, males may overreport coding use to conform to masculine technical identity norms, while females may underreport technical applications to avoid violating gender-typical behaviour expectations. Such response biases would artificially inflate observed gender gaps beyond true behavioural differences, potentially exaggerating the functional stratification we documented.

Lavidas et al. (2022) demonstrate that social desirability effects in student self-reports vary significantly by context, with students exhibiting different response patterns in lecture halls versus laboratory settings where technical competence is more salient. In our study, survey administration through instructor referrals may have activated evaluative concerns, particularly for female STEM students in male-dominated programmes where technical identity is under constant scrutiny. This context-dependent desirability effect could contribute to the apparent "STEM amplification" we observed, where female engineering students may have systematically underreported coding-focused ChatGPT use to manage impressions in technically evaluative contexts. If social desirability operates more strongly in STEM than non-STEM contexts, our observed 33.4-point gap in Applied Sciences could partially reflect measurement artifact rather than true behavioural differences.

Additionally, recall bias may differentially affect responses across task types. Coding tasks are often more salient and memorable than routine text-processing activities (proofreading, summarising), potentially leading males to more readily recall and report technical applications while females more readily recall communicative uses. The 12-task Likert scale format requiring retrospective frequency estimates over unspecified timeframes is particularly susceptible to such retrieval asymmetries. If males disproportionately remember coding instances while females disproportionately remember writing instances, our latent profiles may partially reflect differential recall rather than differential behaviour.

Survey anonymity and online administration mitigate but do not eliminate desirability bias. While participants could not be identified individually, awareness that data would be aggregated by gender and field may have activated group-level identity concerns. Female engineering students, cognisant of stereotypes about women's technical capabilities, may have felt implicitly evaluated as representatives of their gender category, prompting conservative reporting of AI-assisted coding regardless of actual usage.

Methodological recommendations

Future studies should triangulate self-reports with behavioural trace data. ChatGPT API logs, with appropriate consent and anonymisation protocols, would provide objective usage metrics capturing: (1) prompt types submitted (classifiable as technical vs. communicative via natural language processing), (2) session duration and frequency, (3) iterative refinement patterns indicating depth of engagement, and (4) task sequences revealing functional versatility. Such data would isolate true behavioural patterns from self-presentation effects. Mixed-methods approaches combining surveys with qualitative interviews could illuminate how social desirability shapes responses, while experimental designs manipulating survey framing (e.g., "learning strategies" vs. "gender differences in AI use") could test whether identity salience affects reporting patterns. Until such methodological triangulation occurs, our findings should be interpreted as documenting reported usage patterns that may partially reflect perceived appropriateness rather than actual behaviour.

In the Malaysian context specifically, social desirability pressures operate through additional cultural channels that may amplify these effects beyond what Western studies typically report. Malaysia's collectivist orientation means students are acutely sensitive to peer group expectations and face-saving norms (Raza et al., 2024; Raza & Singh, 2024). A female engineering student who uses ChatGPT for coding may be reluctant to report this if she perceives it as drawing attention to a skill gap, while a male student in the same situation may frame identical behaviour as technical resourcefulness. Islamic ethical frameworks present in Malaysian universities may further shape how students report AI-assisted academic work, particularly where institutional messaging conflates AI use with academic dishonesty. Additionally, Malaysia's multilingual environment means that students completing an English-language survey may underreport nuanced or stigmatised behaviours due to language-mediated self-presentation effects, defaulting to socially safe responses when uncertain about phrasing. Together, these culturally specific pressures suggest that observed gender gaps in our data, particularly the coding-use gap, may be inflated relative to true behavioural differences. Future Malaysian studies should consider Bahasa Malaysia survey instruments and mixed-methods designs that allow students to contextualise their responses within local cultural norms.

Cross-sectional design and causal inference

The cross-sectional design prevents causal inference. While gender predicts profile membership, reverse causation remains theoretically possible, though implausible given gender is temporally prior to university AI adoption. More plausibly, unmeasured third variables (prior computing experience, disciplinary socialisation, peer influences) may confound observed associations. Longitudinal designs tracking students from matriculation through graduation would clarify whether gender differences in functional AI appropriation emerge immediately upon ChatGPT exposure or develop gradually through socialisation processes. Panel data would enable fixed-effects models controlling for time-invariant individual heterogeneity, isolating causal effects of field exposure and peer composition on usage pattern evolution.

Sampling limitations and selection bias

Convenience sampling through instructor referrals introduces multiple selection biases. The sample likely overrepresents AI-interested, English-proficient, digitally engaged, urban students from institutions with progressive faculty willing to facilitate research participation. Female non-participants may disproportionately represent AI-averse students, meaning our sample underestimates true gender gaps in adoption while potentially overestimating gaps in functional differentiation among adopters. The exclusion of non-users entirely (inclusion criterion required prior ChatGPT use) prevents analysis of adoption barriers versus appropriation patterns. Additionally, the English-language survey excludes Bahasa Malaysia-primary students, who may exhibit different usage patterns influenced by language-mediated access to AI capabilities.

Probability sampling with rural/urban, socioeconomic, linguistic, and institutional stratification would improve representativeness. Oversampling strategies for underrepresented groups (non-binary students, rural students, Bahasa Malaysia speakers, non-adopters) would enable more comprehensive population inference. Weighting adjustments based on known population demographics could partially correct selection bias in secondary data analysis.

Binary gender operationalisation

Binary gender operationalisation (male/female) excludes non-binary experiences. Five participants (1.1%) selecting "Other" or "Prefer not to say" were excluded from gender-stratified analyses due to insufficient statistical power for stable estimation (post-hoc power <10% for any effect involving this group). This exclusion reflects methodological necessity rather than theoretical dismissal of gender diversity, but it limits our ability to understand how non-binary students navigate gendered technology cultures. Non-binary students may face unique pressures, simultaneously experiencing marginalisation in masculine technical spaces while lacking clear gender-aligned usage scripts that cisgender students follow.

Future research with larger samples (minimum $n \geq 50$ non-binary participants) should examine whether non-binary students: (1) exhibit usage patterns distinct from binary categories, (2) show greater functional versatility due to reduced gender-schema constraints, or (3) experience intensified identity threat leading to avoidance. Qualitative research exploring non-binary students' subjective experiences with AI tools in gendered educational contexts would provide essential insights currently absent from quantitative literature.

Unmeasured mediators and mechanisms

While we proposed four mechanisms driving gendered functional stratification (stereotype threat, peer influence, access to informal learning, gender-identity alignment), our secondary data lack direct measures of these constructs. We cannot empirically test whether observed gender-profile associations operate through self-efficacy differences, stereotype threat activation, peer network composition, or identity protection motives. The mechanisms remain theoretically plausible but empirically unverified.

Future research should directly measure proposed mediators through: (1) validated self-efficacy scales capturing domain-specific confidence (programming self-efficacy, writing self-efficacy), (2) experimental manipulations of stereotype threat (priming gender stereotypes about technical ability vs. neutral primes), (3) social network analysis mapping peer influence on tool adoption, and (4) implicit association tests revealing automatic gender-technology associations. Structural equation modelling with measured mediators would partition total gender effects into direct and indirect pathways, clarifying causal mechanisms. Without such mediation analysis, our mechanistic interpretations remain speculative.

Small cell sizes in interaction effects

Small cell sizes in Gender × Field interactions produce statistically unstable estimates with wide confidence intervals and inadequate statistical power. The Female × Natural Sciences cell ($n = 3$) yields a 90-fold confidence interval range [0.03, 2.71], rendering the estimate scientifically uninformative. Even the more stable Female × Applied Sciences cell ($n = 75$) produces a 6-fold confidence interval range [0.16, 0.95], indicating the true gender gap could range from 8 to 51 percentage points. Post-hoc power analysis reveals 35% power for Natural Sciences and 65% power for Applied Sciences interactions, well below the 80% conventional threshold.

These limitations prevent definitive conclusions about whether STEM contexts amplify gender gaps. Replication with minimum $n \geq 50$ per Gender×Field cell would achieve adequate power. Alternatively, multi-institutional collaborations pooling data across universities could accumulate sufficient cell sizes. Meta-analytic approaches synthesising multiple small studies would provide more stable interaction effect estimates than any single-site study.

Temporal and technological constraints

Data collection (October 2024–February 2025) occurred 24 months post-ChatGPT launch, representing a transitional period when usage patterns may remain fluid. Early adopters in our sample may exhibit different functional orientations than later adopters or non-adopters, limiting generalisability to the full student population. Additionally, rapid AI advancement means ChatGPT's capabilities evolved substantially between November 2022 (GPT-3.5) and our data collection period (GPT-4 widely available). Students' functional appropriation may shift as capabilities expand, plugins emerge, and institutional policies crystallise.

Cohort studies tracking multiple student cohorts across ChatGPT's diffusion curve would distinguish transient adoption dynamics from stable appropriation patterns. If functional stratification diminishes as tools become ubiquitous and normalised, our findings represent early-stage phenomena. Conversely, if stratification persists or intensifies, this suggests durable mechanisms requiring sustained intervention. The rapid pace of AI advancement may render specific findings obsolete even before publication, though theoretical insights about gendered technology appropriation likely retain validity across successive AI generations.

Generalisation and cultural specificity

Findings from Malaysia, a middle-income, Muslim-majority, multiethnic, multilingual nation with robust digital infrastructure yet pronounced occupational gender segregation, may not generalise to Western or culturally distinct contexts. Malaysia's unique configuration of educational gender parity (60% female enrolment) coexisting with STEM under-representation (22-28% female in engineering/computing) creates specific conditions that may intensify functional stratification. Collectivist cultural orientations emphasising conformity and social harmony may amplify peer influence mechanisms, while Islamic frameworks shaping gender roles may influence perceived appropriateness of technical versus communicative tool uses.

Cross-cultural replications in Western contexts (individualist, higher gender equality indices) and other non-Western contexts (Sub-Saharan Africa, Latin America) would establish boundary conditions. If gendered functional differentiation persists across diverse cultural settings, this would support universal mechanisms (stereotype threat, identity alignment). If patterns attenuate in gender-equitable societies (Nordic countries) or intensify in patriarchal contexts (Middle East/North Africa), this would confirm cultural moderation. Comparative research would clarify whether third-level digital divides represent global phenomena requiring coordinated international intervention or culturally specific challenges demanding localised solutions.

Strengths amid limitations

Despite these limitations, the study provides valuable contributions. The person-centred latent profile approach captures heterogeneity in functional use that variable-centred analyses overlook. The 12-item task-specific measurement offers fine-grained resolution of third-level digital divide patterns. Seven covariates and interaction tests enhance model precision, while six robustness checks (K-means validation, cross-validation, outlier exclusion, imputation comparison, measurement invariance, bootstrap confidence intervals) confirm analytical stability. The Malaysian setting extends gender and AI literature beyond Western samples, testing theoretical generalisability. Data collection timing (24 months post-launch) captures a critical equilibration period when AI integration had normalised in hybrid learning environments, providing ecological validity. Large effect sizes (η^2 = .18-.33) for profile-outcome associations demonstrate practical significance despite statistical limitations in interaction effects. These strengths ensure findings contribute meaningfully to literature despite acknowledged constraints.

CONCLUSION

This study identified four behavioural profiles among Malaysian university students, Selective Users (14.9%), Moderate Adopters (31.8%, coding-focused), Academic Enthusiasts (33.0%, text-focused), and Comprehensive Users (20.3%), revealing that gender differences operate at the level of functional differentiation rather than basic access. While students show near-universal ChatGPT awareness (85% adoption) with no gender gaps in minimal engagement ($p > .05$), systematic differences emerge in which applications students use: males concentrate in coding-focused use (OR = 0.48 for females, $p = .007$), females in text-focused use (OR = 1.89, $p = .020$), with exploratory analyses suggesting potentially larger gaps in STEM fields (33.4 points in Applied Sciences) versus Social Sciences (12.7 points), though small sample sizes ($n=3$ for female Natural Sciences students) prevent definitive conclusions. These patterns broadly align with prior research on gendered technology appropriation, though the specific STEM amplification pattern requires replication with adequately powered samples.

The Comprehensive Users profile's gender parity demonstrates equitable AI literacy is achievable, yet only 20% reach this profile, requiring curriculum mandates promoting counter-stereotypical exploration, faculty development, culture change, and policy accountability. For Malaysia, having achieved quantitative gender parity (60% female enrolment), qualitative equity remains elusive. While our data show clear gender differences in functional AI appropriation, the extent to which these educational patterns prefigure occupational segregation requires longitudinal investigation tracking students into the workforce. Importantly, the gender gap was not moderated by institution type or learning format, suggesting mechanisms driving the third-level digital divide are rooted in broader socio-cultural contexts rather than institutional policies, underscoring that preventing occupational segregation requires systemic intervention addressing functional stratification, a new analytical layer within digital equity research demanding attention as generative AI threatens to erode hard-won gender parity gains.

Acknowledgement

The authors wish to thank the Global ChatGPT Student Survey team led by Aristovnik et al. (2025) for making the dataset publicly available under a CC BY 4.0 licence, and the Malaysian students whose responses form the basis of this study.

Funding

This research received no external funding.

Ethical statement

This study is based on secondary analysis of a publicly available, fully anonymised dataset (Aristovnik et al., 2025) published under a CC BY 4.0 licence via Mendeley Data (https://doi.org/10.17632/nv2343nwsb.2). The parent study received ethical approval from institutional review committees at all participating universities in accordance with the Declaration of Helsinki. No additional ethical approval was required for this secondary analysis, as the data were de-identified prior to access, no contact with participants occurred, and the analyses fell within the scope of the original consent framework, which explicitly permitted demographic subgroup investigations.

Competing interests

The authors declare no competing interests.

Author contributions

Fahd Ali Raza: conceptualisation, methodology, formal analysis, data curation, writing – original draft, writing – review and editing. Abtar Darshan Singh: supervision, writing – review and editing. Rabia Anwar: writing – review and editing. Jonathan Jeevan Srinivas Kovilpillai: writing – review and editing. Analisa Binti Hamdan: writing – review and editing. Fumiko Konno: writing – review and editing. Vaikunthan Rajaratnam: writing – review and editing. Murali Raman: supervision, writing – review and editing. Husna Hafiza Razami: writing – review and editing.

Data availability

The dataset analysed in this study is publicly available via Mendeley Data: Aristovnik, A. et al. (2025). Higher education students' evolving perceptions of ChatGPT: Global survey data from the academic year 2024–2025 (Version 2). https://doi.org/10.17632/nv2343nwsb.2

AI disclosure

No artificial intelligence tools were used in the writing, analysis, or preparation of this manuscript.

Biographical sketch

Dr. Fahd Ali Raza is a Lecturer and Digital Learning Designer at the Digital Learning Hub, Asia Pacific University of Technology and Innovation, Kuala Lumpur, Malaysia. With more than seventeen years of international academic experience, he works at the intersection of digital learning design and educational innovation. His research focuses on generative AI in education, gender equity in STEM leadership, and technology enabled learning environments. His work explores how emerging technologies reshape teaching practice, digital capability development, and inclusive innovation in higher education.

Prof. Abtar Darshan Singh is Professor of Innovative Digital Learning and Director of the Digital Learning Hub at the Asia Pacific University of Technology and Innovation, Malaysia. He also holds the UNESCO Chair in Harnessing Innovative Technologies to Enhance Quality Teaching and Learning. His work focuses on artificial intelligence in education, digital pedagogy, and open and distance learning. He leads regional initiatives on technology enabled education, educator capacity development, and international collaborations aimed at advancing digital transformation in higher education.

Ts. Jonathan Jeevan Srinivas Kovilpillai is affiliated with the Digital Learning Hub at the Asia Pacific University of Technology and Innovation, Kuala Lumpur, Malaysia. His work focuses on digital learning innovation, technology enabled teaching practices, and institutional digital transformation. He contributes to initiatives that support the development of AI supported learning environments and scalable digital learning solutions. His professional interests include instructional technology implementation, learning analytics, and digital education strategy within higher education institutions.

Ms. Analisa Hamdan is an academic and researcher at the Asia Pacific University of Technology and Innovation, Malaysia. Her work focuses on digital pedagogy, educational innovation, and technology enhanced learning environments. She has contributed to research on digital pedagogical tools, digital competency development, and the integration of emerging technologies into teaching and learning. Her research interests include student engagement in digital learning environments and the use of educational technologies to improve learning experiences in higher education.

Dr. Vaikunthan Rajaratnam is a Senior Consultant Hand Surgeon at Khoo Teck Puat Hospital in Singapore and a Clinical Senior Lecturer at the National University of Singapore School of Medicine. His work spans clinical medicine, medical education, and instructional design. His research interests include hand and peripheral nerve surgery, surgical education, and the application of technology and digital learning approaches in healthcare training and professional education.

Prof. Murali Raman is Deputy Vice Chancellor for Academic Development and Strategy at the Asia Pacific University of Technology and Innovation, Malaysia. A professor of technology management and innovation, his research focuses on digital transformation, design thinking, and innovation ecosystems. His work examines how technology adoption and human centred design approaches drive organizational change, entrepreneurship, and digital innovation across education and industry.

Dr. Fumiko Konno is a Lecturer and Senior Digital Learning Designer at the Digital Learning Hub, Asia Pacific University of Technology and Innovation, Malaysia. She holds a PhD in Educational Informatics and has experience in faculty development and instructional design. Her research focuses on digital pedagogy, instructional design, teacher professional development, and technology enhanced learning. Her work supports academics in redesigning courses and learning experiences using digital and AI supported educational technologies.

Dr. Husna Hafiza Razami is an academic and researcher at the Asia Pacific University of Technology and Innovation, Malaysia. Her research focuses on educational technology, digital learning adoption, and technology acceptance in higher education. She has published work on MOOCs, digital educational games, and online learning environments. Her research examines how emerging learning technologies influence student engagement, learning effectiveness, and digital learning ecosystems in higher education.

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 do not necessarily reflect the views of Lectito Publications and/or the editor(s). Lectito Publications and/or the editor(s) disclaim responsibility for any injury to persons or property resulting from any ideas, methods, instructions, or products referred to in the content.

REFERENCES

Acar, E., Yigit, F., & Deiri, Y. (2025). A focused review of artificial intelligence in education: Evolution and challenges. *Journal of Interdisciplinary Research in Artificial Intelligence and Society, 1*(1), 3. https://doi.org/10.20897/jirais/17640

Ahn, J., Kim, J., & Sung, Y. (2022). The effect of gender stereotypes on artificial intelligence recommendations. *Journal of Business Research, 141,* 50–59. https://doi.org/10.1016/j.jbusres.2021.12.007

Al-Adwan, A., Li, N., Al-Adwan, A., Abbasi, G., Albelbisi, N., & Habibi, A. (2023). Extending the technology acceptance model (TAM) to predict university students' intentions to use metaverse-based learning platforms. *Education and Information Technologies, 28*. https://doi.org/10.1007/s10639-023-11816-3

Alam, A. (2022). Psychological, sociocultural, and biological elucidations for gender gap in STEM education: A call for translation of research into evidence-based interventions. *Atlantis Highlights in Social Sciences, Education and Humanities*. https://doi.org/10.2991/ahsseh.k.220105.012

Alier, M., García-Peñalvo, F., & D. Camba, J. (2024). Generative artificial intelligence in education: From deceptive to disruptive. *International Journal of Interactive Multimedia and Artificial Intelligence, 8*(5). https://doi.org/10.9781/ijimai.2024.02.011

Aristovnik, A., Ravšelj, D., Keržič, D., Tomaževič, N., Umek, L., Brezovar, N., Kessler, S. H., Schäfer, M., & Al, E. (2025). *Higher education students' evolving perceptions of ChatGPT: Global survey data from the academic year 2024–2025*. (Version 2) [Data set]. Mendeley Data. https://doi.org/10.17632/nv2343nwsb.2

Austin, P. C., White, I. R., Lee, D. S., & van Buuren, S. (2021). Missing data in clinical research: A tutorial on multiple imputation. *Canadian Journal of Cardiology, 37*(9), 1322–1331. https://doi.org/10.1016/j.cjca.2020.11.010

Awoniyi, F. C., & Jokotagba, S. T. (2025). Parental influence, career counselling, and role models in girls' STEM career choices: Evidence from three public universities in Ghana. *European Journal of STEM Education, 10*(1), Article 05. https://doi.org/10.20897/ejsteme/16432

Banton, C., Garza, J. A., Goods, A., Jones, T., & Langford, L. (2024). Examining the retention and exodus of women in technology. *American Journal of Qualitative Research, 8*(4), 132-150. https://doi.org/10.29333/ajqr/15215

Basch, C., Hillyer, G., Gold, B., & Yousaf, H. (2025). Understanding artificial intelligence knowledge and usage among college students: Insights from a survey on classroom, coursework, and personal applications. *Review of Artificial Intelligence in Education, 6*(i), e037. https://doi.org/10.37497/rev.artif.intell.educ.v6ii.37

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2024). *Generative AI can harm learning*. (SSRN Scholarly Paper No. 4895486). Social Science Research Network (SSRN). https://doi.org/10.2139/ssrn.4895486

Bicer, A., Lee, Y., & Perihan, C. (2020). Inclusive STEM high school factors influencing ethnic minority students' STEM preparation. *Journal of Ethnic and Cultural Studies, 7*(2), 147–172. https://doi.org/ 10.29333/ejecs/384

Biswas, S. (2023). ChatGPT and the future of medical writing. *Radiology, 307,* 223312. https://doi.org/10.1148/radiol.223312

Blazek, K., van Zwieten, A., Saglimbene, V., & Teixeira-Pinto, A. (2021). A practical guide to multiple imputation of missing data in nephrology. *Kidney International, 99*. https://doi.org/10.1016/j.kint.2020.07.035

Cheok, C. K. (2024). *Higher education in Malaysia*. In L. P. Symaco (Ed.), *Higher education in Southeast Asia, 49,* 49–65. Emerald Publishing Limited. https://doi.org/10.1108/S1479-367920240000049005

Clarke, N., Mondisa, J.-L., Packard, B. W.-L., Queener Schemanske, C., Tuladhar, A., & Gosha, K. (2024). Examining the role of computing identity in the computing experiences of women and racially minoritized undergraduates: A literature review. *Journal of Computing in Higher Education, 36*(3), 773–812. https://doi.org/10.1007/s12528-023-09375-0

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

Depaoli, S., Jia, F., & Visser, M. (2024). Addressing missing data in latent class analysis when using a three-step estimation approach. *Structural Equation Modeling: A Multidisciplinary Journal, 32,* 1–17. https://doi.org/10.1080/10705511.2024.2410240

Dergaa, I., Chamari, K., Glenn, J. M., Ben Aissa, M., Guelmami, N., & Ben Saad, H. (2023). Towards responsible research: Examining the need for preprint policy reassessment in the era of artificial intelligence. *EXCLI Journal, 22,* 686–689. https://doi.org/10.17179/excli2023-6324

Devitt, M. (2023). *Biological essentialism*. Oxford University Press

Dyczek, B. (2025). Socio-cultural framework for Implementing AI Technologies into Education in Malaysia. *SGS - Humanities & Management, 1*(1). https://spast.org/index.php/sgshm/article/view/5258

Elhadary, Y., & Samat, N. (2023). Addressing gender disparity in public universities of Malaysia: Challenges and achievements. *Journal of Higher Education Theory and Practice, 23*(14). https://doi.org/10.33423/ jhetp.v23i14.6390

FakhrHosseini, S., Chan, K., Lee, C., Jeon, M., Son, H., Rudnik, J., & Coughlin, J. (2024). User adoption of intelligent environments: A review of technology adoption models, challenges, and prospects. *International Journal of Human–Computer Interaction, 40*(4), 986–998. https://doi.org/10.1080/10447318.2022.2118851

Gill, S. S., Xu, M., Patros, P., Wu, H., Kaur, R., Kaur, K., Fuller, S., Singh, M., Arora, P., Parlikad, A. K., Stankovski, V., Abraham, A., Ghosh, S. K., Lutfiyya, H., Kanhere, S. S., Bahsoon, R., Rana, O., Dustdar, S., Sakellariou, R., … Buyya, R. (2024). Transformative effects of ChatGPT on modern education: Emerging era of AI chatbots. *Internet of Things and Cyber-Physical Systems, 4,* 19–23. https://doi.org/10.1016/ j.iotcps.2023.06.002

Halaweh, M. (2023). ChatGPT in education: Strategies for responsible implementation. *Contemporary Educational Technology, 15*. https://doi.org/10.30935/cedtech/13036

Hassan, S. (2024). *Shaping the future of the internet we want*. UUM Press.

Hassan, S. M. M., Marzuki, A. S. M., Amiruddin, M. F., Bujang, S., Samingan, A. K., Salleh, M. K., Sulaiman, A., Nor, M. H. M., Ismail, M. A., Ishak, M. S., & Abidin, K. B. Z. (2021). Bridging the digital divide in Malaysia using fixed wireless access. *Proceeding - 2021 26th IEEE Asia-Pacific Conference on Communications, APCC 2021,* 74–78. https://doi.org/10.1109/APCC49754.2021.9609825

Hendrian, S. (2025). Digital inequality in the age of AI: Inequality of access and technology literacy between urban and rural communities in Southeast Asia. *Insight: International Journal of Social Research, 3*(3). https://doi.org/10.59888/insight.v3i3.62

Hofer, M., & Hargittai, E. (2024). Online social engagement, depression, and anxiety among older adults. *New Media & Society, 26*(1), 113–130. https://doi.org/10.1177/14614448211054377

Homeister, L. (2024). *Application and comparison of missing data methods for factor analysis and multinomial logistic regression* (Doctoral dissertation, University of Washington). ProQuest Dissertations & Theses Global.

Huynh, Q. Q. (2024). A Literature review on EFL students' integration of Al-based tools in English learning and the perceived impact on autonomy. *International Journal of AI in Language Education, 1*(2), 30–51. https://doi.org/10.54855/ijaile.24123

Işıklı, Ş., & Fazlıoğlu, E. F. (2026). Technological effects on gender studies: An intersectional perspective. *Feminist Encounters: A Journal of Critical Studies in Culture and Politics, 10*(1), Article 23. https://doi.org/10.20897/femenc/17998

Jamaluddin, F., Jamaluddin, A. H., Jamaluddin, F., & Jamaluddin, F. (2025). *Malaysia's AI-Driven education landscape: Policies, applications, and comparative insights for a digital future* (No. arXiv:2509.21858). arXiv. https://doi.org/10.48550/arXiv.2509.21858

Karakose, T., Tülübaş, T., & Papadakis, S. (2023). The scientific evolution of social justice leadership in education: Structural and longitudinal analysis of the existing knowledge base, 2003–2022. *Frontiers in Education, 8,* 1139648. https://doi.org/10.3389/feduc.2023.1139648

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Katz, N. (2026). Digital inequality in context: A socio-technical analysis of Arab students' remote learning in Israel. *Asia Pacific Journal of Education and Society, 14*(1), 6. https://doi.org/10.20897/apjes/17975

Kaur, D., & Kaur, R. (2022). Elucidating the role of gender differences via TAM in e-recruitment adoption in India: A multi-group analysis using MICOM. *The Bottom Line, 35*(2–3), 115–136. https://doi.org/ 10.1108/BL-11-2021-0104

Kelly, A., Sullivan, M., & Strampel, K. (2025). *Teaching generative AI in higher education: Strategies, implications, and reflective practices*. In *teaching and learning in the age of generative AI: Evidence-based approaches to pedagogy, ethics, and beyond* (pp. 213–234). Routledge. https://doi.org/10.4324/9781032688602-14

Klinger, U., & Svensson, J. (2023). *The power of code: Women and the making of the digital world*. In A. Schiffrin, K. Koc-Michalska, & M. Ferrier (Eds.), *Women in the digital world* (pp. 84–100). Routledge. https://doi.org/10.4324/9781003375937-6

Kuah, A. T. H., Kim, C., & Le Queux, S. (2020). Multiculturalism in Singapore and Malaysia: Approaches and outcomes. *Equality, Diversity and Inclusion: An International Journal*. https://doi.org/10.1108/EDI-05-2019-0148

Lagesen, V. (2008). A cyberfeminist utopia? Perceptions of gender and computer science among Malaysian women computer science students and faculty. *Science Technology & Human Values, 33,* 5–27. https://doi.org/10.1177/0162243907306192

Lalumera, E. (2024). An overview on trust and trustworthiness: Individual and institutional dimensions. *Philosophical Psychology, 37*(1), 1–17. https://doi.org/10.1080/09515089.2024.2301860

Lamberti, G., López-Sintas, J., & Lopez Belbeze, P. (2023). The impact of internet use on leisure: Gender and age heterogeneity in young people. *Journal of Leisure Research, 54*(5), 581–601. https://doi.org/10.1080/00222216.2023.2193178

Lampropoulos, G., & Papadakis, S. (2025). *The educational value of artificial intelligence and social robots. In Artificial intelligence and social robots in education* (pp. 3–18). Springer. https://doi.org/10.1007/978-3-031-82915-4_1

Lancaster, T. (2023). Artificial intelligence, text generation tools and ChatGPT – does digital watermarking offer a solution? *International Journal for Educational Integrity, 19*(1), 10. https://doi.org/10.1007/s40979-023-00131-6

Lau, S., & Guo, P. (2023). From "Ban it till we understand it" to "Resistance is futile": How university programming instructors plan to adapt as more students use AI code generation and explanation tools such as ChatGPT and GitHub Copilot. *ICER 2023 - Proceedings of the 2023 ACM Conference on International Computing Education Research, 1*(16), 106–121. https://doi.org/10.1145/3568813.3600138

Lavidas, K., Papadakis, S., Manesis, D., Grigoriadou, A. S., & Gialamas, V. (2022). The effects of social desirability on students' self-reports in two social contexts: Lectures vs. lectures and lab classes. *Information, 13*(10), 491. https://doi.org/10.3390/info13100491

Li, X. (2023). Strategic flexibility in a male-dominated occupation: Women software engineers in China. *Journal of Gender Studies, 32*(4), 330–342. https://doi.org/10.1080/09589236.2021.2006615

Lin, Z. (2023). Why and how to embrace AI such as ChatGPT in your academic life. *Royal Society Open Science, 10*(8). https://doi.org/10.1098/RSOS.230658

Low, M. P., Wut, T. M., Lau, T. C., & Tong, W. (2025). The interplay of self-efficacy, artificial intelligence literacy and lifelong learning for career resilience among older employees: A comparison study between China and Malaysia. *Current Psychology, 44*(9), 7879–7896. https://doi.org/10.1007/s12144-025-07434-6

Lydersen, S. (2022). Multiple imputation of missing data. *Tidsskrift for Den norske legeforening, 142.* https://doi.org/10.4045/tidsskr.21.0772

Mastam, M. (2024). Bridging the digital divide in Malaysia: Enhancing digital literacy for inclusive students in educational systems. *Asia Pacific Journal of Youth Studies, 15,* 128–150. https://doi.org/10.56390/apjys2024.15.2.128

Møgelvang, A., Bjelland, C., Grassini, S., & Ludvigsen, K. (2024). Gender differences in the use of generative artificial intelligence chatbots in higher education: Characteristics and consequences. *Education Sciences, 14.* https://doi.org/10.3390/educsci14121363

Moore, E. W. G., & Quartiroli, A. (2025). Representing subpopulations with latent profile analysis: A non-technical introduction using exercisers' goal orientation adoption profiles. *Journal of Behavioral Medicine.* https://doi.org/10.1007/s10865-025-00596-5

Mustapha, Y., Hilmi, M. F., Faiz, M., & Saad, S. M. (2022). Prospect and challenges of blended learning in Malaysia: A systematic literature review. *Asian Journal of Research in Education and Social Sciences, 4*(3), 51–61. https://doi.org/10.55057/ajress.2022.4.3.5

Nohuddin, P. N. E., AbdKadir, Z., & Noordin, N. A. (2025). *A framework for bridging the digital divide: Improving connectivity and opportunities in rural Malaysia*. In L. O. Yesufu & P. N. E. Nohuddin (Eds.), *Technology for societal transformation* (pp. 215–233). Springer Nature Singapore. https://doi.org/10.1007/978-981-96-1721-0_14

Nylund-Gibson, K., Garber, A. C., Carter, D. B., Chan, M., Arch, D. A. N., Simon, O., Whaling, K., Tartt, E., & Lawrie, S. I. (2023). Ten frequently asked questions about latent transition analysis. *Psychological Methods, 28*(2), 284–300. https://doi.org/10.1037/met0000486

Otis, N. G., Delecourt, S., Cranney, K., & Koning, R. (2024). *Global evidence on gender gaps and generative AI*. Harvard Business School. https://www.hbs.edu/ris/Publication%20Files/25-023_8ee1f38f-d949-4b49-80c8-c7a736f2c27b.pdf

Papadakis, S., Semerikov, S. O., Striuk, A., Danylchuk, H., et al. (2024, May). *Embracing digital innovation and cloud technologies for transformative learning experiences*. In *Proceedings of the 11th workshop on cloud technologies in education (CTE 2023),* Kryvyi Rih, Ukraine, December 22, 2023.

Patras, Y. E., Hidayat, R., Maksum, A., & Nurhasanah, N. (2022). Understanding multiculturalism education from Indonesia, Singapore, Malaysia, and Thailand. *Kelola: Jurnal Manajemen Pendidikan, 9*(2), 125–135. https://doi.org/10.24246/j.jk.2022.v9.i2.p125-135

Pérez, H. E. (2024). Your p-values are significant (or not), so what … now what? *Seed Science Research, 34*(2), 86–89. https://doi.org/10.1017/S0960258524000035

Rahim, Z. A., & Iqbal, M. S. (2025). Bridging digital gaps: A framework to enhance digital economy competences in Malaysia. *Journal of Advanced Research in Applied Sciences and Engineering Technology, 64,* 219–238. https://doi.org/10.37934/araset.64.2.219238

Rahman, M. M., & Watanobe, Y. (2023). ChatGPT for education and research: Opportunities, threats, and strategies. *Applied Sciences (Switzerland), 13*(9). https://doi.org/10.3390/APP13095783

Rahman, N., & Halim, L. (2022). STEM career interest: The effect of gender. *Creative Education, 13,* 2530–2543. https://doi.org/10.4236/ce.2022.138160

Raza, F. A., & Singh, A. D. (2024). Unveiling the missing link: Women in STEM leadership – A comprehensive review. *International Journal of Advanced Business Studies, 3*(Special Issue), 15–27. https://doi.org/10.59857/IJABS.1109

Raza, F. A., Singh, A. D., Kovilpillai, J. J. S., Hamdan, A., & Rajaratnam, V. (2025). Safeguarding integrity in AI-enhanced education: Stakeholder perspectives on accuracy, validity, and ethics in ASEAN. *European Journal of STEM Education, 10*(1), 22. https://doi.org/10.20897/EJSTEME/17307

Raza, F., Singh, A., Kovilpillai, J., & Hamdan, A. (2024). Gender gaps and convergence: ASEAN stakeholder perspectives on artificial intelligence in education. *Journal of Research in Gender Studies, 14*(2), 34. https://doi.org/10.22381/JRGS14220242

Rizvi, I., Bose, C., & Tripathi, N. (2025). *Transforming education: Adaptive learning, AI, and online platforms for personalization*. In L. O. Yesufu & P. N. E. Nohuddin (Eds.), *Technology for societal transformation: exploring the intersection of information technology and societal development* (pp. 45–62). Springer Nature. https://doi.org/10.1007/978-981-96-1721-0_4

Saadat, Z., & Sultana, A. M. (2023). Understanding gender disparity: Factors affecting higher education self-efficacy of students in Malaysia. *Journal of Science and Technology Policy Management.* https://doi.org/10.1108/ JSTPM-10-2022-0165

Seong, M. C., Fauzi, M. F., Juhari, S. N., Aliman, N., Rajoo, M., Abdul Wahab, N. M., & Magiman, M. M. (2022). Blended learning practices in Malaysia higher education: A review. *Mathematical Statistician and Engineering Applications, 71*(3), 1637–1652. https://doi.org/10.17762/msea.v71i3.1497

Shaari, R., Kamarudin, D., Ju, S.-Y., & Zakaria, H. (2022). Effects of leadership types on job satisfaction among Malaysian higher education institutions. *Asian Journal of Instruction, 10*(1), 54-70. https://doi.org/10.47215/aji.1020324

Sharipova, G. (2022). A brief history of multiculturalism in Malaysia. *Russia and the Moslem World, 3*(317), 90–100. https://doi.org/10.31249/rmw/2022.03.09

Shrivastava, A. (2024). *Transformative landscape of ChatGPT in higher education: Unveiling the academic discourse. Global knowledge, memory and communication.* Advance online publication. https://doi.org/10.1108/GKMC-02-2024-0068

Sintas, J. L., Belbeze, M. P. L., & Lamberti, G. (2023). Socially patterned strategic complementarity between offline leisure activities and internet practices among young people. *Leisure Sciences, 45*(1), 24–45. https://doi.org/10.1080/01490400.2020.1782289

Skjuve, M., Brandtzaeg, P. B., & Følstad, A. (2024). Why do people use ChatGPT? Exploring user motivations for generative conversational AI. *First Monday, 29*(1). https://doi.org/10.5210/FM.V29I1.13541

Soon Tan, C., Zakuan, N., & Ismail Abd Aziz, M. (2022). Recent trends of blended learning and flipped classroom in Malaysia. *Arab World English Journal (AWEJ), 2nd Special Issue on COVID-19 Challenges.* https://doi.org/10.2139/ssrn.4036764

Strzelecki, A. (2024). Students' acceptance of ChatGPT in higher education: An extended unified theory of acceptance and use of technology. *Innovative Higher Education, 49*(2), 223–245. https://doi.org/10.1007/ s10755-023-09686-1

Subramaniam, L. (2023). *Digital divide among Malaysian tertiary students and its impact on online learning during the COVID-19 pandemic* (MPhil dissertation, Universiti Tunku Abdul Rahman). http://eprints.utar.edu.my/5910/1/Latha_(2002951).pdf

Taib, S. M., Mohd Nazif, N., Ariffin, A., & Saman, N. (2025). Towards a responsible AI governance framework: Lessons from policy implementation in Malaysia. *Zorig Melong: A Technical Journal of Science, Engineering and Technology, 8,* 223–230. https://doi.org/10.17102/zmv8.i2.025

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments, 10*(1), 15. https://doi.org/10.1186/s40561-023-00237-x

Uğraş, H., Uğraş, M., Papadakis, S., & Kalogiannakis, M. (2024). ChatGPT-supported education in primary schools: The potential of ChatGPT for sustainable practices. *Sustainability, 16*(22), 9855. https://doi.org/10.3390/su16229855

Unger, R. (2020). *Representations: Social constructions of gender*. Routledge.

Van Deursen, A. J. A. M., Van Der Zeeuw, A., De Boer, P., Jansen, G., & Van Rompay, T. (2021). Digital inequalities in the internet of things: Differences in attitudes, material access, skills, and usage. *Information, Communication & Society, 24*(2), 258–276. https://doi.org/10.1080/1369118X.2019.1646777

Van Deursen, A. J., & Van Dijk, J. A. (2023). IQ and digital inequality: An empirical investigation. *New Media & Society, 25*(6), 1248–1270. https://doi.org/10.1177/14614448211024012

Venkatesh, V., Speier, C., & Morris, M. G. (2002). User acceptance enablers in individual decision making about technology: Toward an integrated model. *Decision Sciences, 33*(2), 297–316. https://doi.org/10.1111/j.1540-5915.2002.tb01646.x

Walshe, P., Commins, A., McDonnell, C., & Kelly, B. G. (2025). 'STEAM from the Start': Proposing a conceptual framework for the development and implementation of a STEAM training intervention for early childhood educators. *European Journal of STEM Education, 10*(1), 11. https://doi.org/10.20897/ejsteme/17156

Webb, R. (2026). Post pandemic English language teacher development: A global perspective. *European Journal of Education & Language Review, 2*(1), Article 1. https://doi.org/10.20897/ejelr/17719

Wickneswary, N., Rahman, A., Rahman, A., Haque, R., Udang, L., Osman, Z., Alainati, S., Al-Hunaiyyan, A., Bin, M., Mohd Isa, A. M., & Ramasamy, G. (2024). Factors influencing college students' educational enrolment choice in private higher education institutions in Klang Valley, Malaysia. *Kurdish Studies, 12*(2), 3674–3693. https://doi.org/10.58262/ks.v12i2.274

Wölfel, M., Shirzad, M. B., Reich, A., & Anderer, K. (2024). Knowledge-based and generative-AI-Driven pedagogical conversational agents: A comparative study of Grice's cooperative principles and trust. *Big Data and Cognitive Computing, 8*(1), 2. https://doi.org/10.3390/bdcc8010002

Woods, A. D., Gerasimova, D., Van Dusen, B., Nissen, J., Bainter, S., Uzdavines, A., Davis-Kean, P. E., Halvorson, M., King, K. M., Logan, J. A. R., Xu, M., Vasilev, M. R., Clay, J. M., Moreau, D., Joyal-Desmarais, K., Cruz, R. A., Brown, D. M. Y., Schmidt, K., & Elsherif, M. M. (2024). Best practices for addressing missing data through multiple imputation. *Infant and Child Development, 33*(1), e2407. https://doi.org/ 10.1002/icd.2407

Xing, W., Nixon, N., Crossley, S., et al. (2025). The use of large language models in education. *International Journal of Artificial Intelligence in Education, 35,* 439–443. https://doi.org/10.1007/s40593-025-00457-x

Yadegari, M., Mohammadi, S., & Masoumi, A. H. (2024). Technology adoption: An analysis of the major models and theories. *Technology Analysis & Strategic Management, 36*(6), 1096–1110. https://doi.org/10.1080/09537325.2022.2071255

Yan, Y., & Liu, H. (2025). Ethical framework for AI education based on large language models. *Education and Information Technologies, 30,* 10891–10909. https://doi.org/10.1007/s10639-024-13241-6

Yang, J., & Zhang, M. (2023). Beyond structural inequality: A socio-technical approach to the digital divide in the platform environment. *Humanities and Social Sciences Communications, 10,* 813. https://doi.org/10.1057/s41599-023-02326-1

Zhou, A. (2021). *Gender codes: Exploring Malaysia's gender parity in computer science* (Doctoral dissertation). Harvard University. https://dash.harvard.edu/server/api/core/bitstreams/b821900d-4902-480a-b4b1-41915f6f700e/content

Appendix A: Technical tables

Table A1

Complete latent profile analysis model comparison statistics

Model	Log-likelihood	AIC	BIC	Entropy	AvePP Profile 1	AvePP Profile 2	AvePP Profile 3	AvePP Profile 4	AvePP Profile 5	AvePP Profile 6	BLRT p-value
2-Profile	-7,124.3	14,296.6	14,382.1	0.812	0.94	0.89	—	—	—	—	<.001
3-Profile	-6,987.5	14,047.0	14,157.8	0.841	0.92	0.87	0.91	—	—	—	<.001
4-Profile	-6,892.1	13,880.2	13,941.5	0.871	0.94	0.89	0.92	0.90	—	—	<.001
5-Profile	-6,831.4	13,782.8	13,919.4	0.868	0.93	0.88	0.91	0.86	0.89	—	.082
6-Profile	-6,789.7	13,723.4	13,935.3	0.854	0.92	0.87	0.89	0.84	0.86	0.88	.194

Note: AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; AvePP = Average Posterior Probability of profile membership (diagonal values from classification table); BLRT = Bootstrap Likelihood Ratio Test comparing k-profile model to (k-1)-profile model. Bold row indicates selected model based on lowest BIC, high entropy (>.80), all AvePP >.80, significant BLRT, and theoretical interpretability. The 5-profile and 6-profile models showed diminishing returns with non-significant BLRT and profile fragmentation (two profiles differed by <0.3 scale points on all indicators).

Table A4

Multinomial logistic regression full model results (n = 438)

Predictor	Comparison: Selective vs. Comprehensive	Comparison: Moderate vs. Comprehensive	Comparison: Academic vs. Comprehensive
	OR [95% CI]	p	OR [95% CI]
Intercept	0.18 [0.02, 1.64]	.126	0.31 [0.05, 1.89]
Gender (Female)	1.03 [0.54, 1.96]	.928	0.48 [0.28, 0.82]
Field: Applied Sciences	1.24 [0.58, 2.65]	.581	2.18 [1.14, 4.17]
Field: Arts/Humanities	0.89 [0.32, 2.47]	.820	1.12 [0.45, 2.79]
Field: Natural Sciences	1.47 [0.38, 5.68]	.577	1.85 [0.54, 6.35]
Academic Level	0.94 [0.73, 1.21]	.632	0.88 [0.71, 1.09]
Institution (Private)	1.18 [0.61, 2.28]	.624	1.05 [0.60, 1.84]
Modality: Fully Online	1.32 [0.52, 3.35]	.561	1.21 [0.54, 2.71]
Modality: Face-to-Face	0.87 [0.43, 1.76]	.695	0.92 [0.51, 1.66]
Age	1.02 [0.93, 1.12]	.694	1.00 [0.93, 1.08]
Usage Intensity	0.31 [0.21, 0.46]	<.001	0.54 [0.40, 0.73]
Experience	0.68 [0.45, 1.03]	.067	0.81 [0.59, 1.11]

Note: Reference categories: Gender = Male, Field = Social Sciences, Institution = Public, Modality = Blended/Hybrid, Profile = Comprehensive Users. OR = Odds Ratio; 95% CI = 95% Confidence Interval. Bold values indicate p < .05. Model fit: Nagelkerke pseudo-R^2 = .238; AIC = 2,140.6; Likelihood ratio $\chi^2(36)$ = 187.4, p < .001. Usage Intensity shows strongest effects: higher intensity reduces odds of being in any profile vs. Comprehensive (all ORs < 0.55, p < .001), validating Comprehensive as highest-engagement group. Gender effects remain significant after controlling all covariates: females have 52% lower odds of Moderate membership (OR = 0.48) and 89% higher odds of Academic membership (OR = 1.89) relative to males.

Table A5

Gender × field interaction model results (selected comparisons)

Predictor	Moderate vs. Comprehensive	Academic vs. Comprehensive
	OR [95% CI]	p
Gender (Female)	0.73 [0.35, 1.52]	.398
Field: Applied Sciences	3.87 [1.64, 9.14]	.002
Gender × Applied Sciences	0.39 [0.16, 0.96]	.038
Field: Arts/Humanities	1.24 [0.41, 3.76]	.702
Gender × Arts/Humanities	0.68 [0.17, 2.71]	.581
Field: Natural Sciences	2.15 [0.48, 9.61]	.318
Gender × Natural Sciences	0.54 [0.08, 3.64]	.524

Note: Reference categories: Gender = Male, Field = Social Sciences. Model includes all covariates from Table A4 (not shown for brevity). Bold values indicate p < .05. The significant Gender × Applied Sciences interaction (OR = 0.39, p = .038) for Moderate vs. Comprehensive comparison indicates that the female disadvantage in Moderate profile membership is 61% smaller (more negative) in Applied Sciences than Social Sciences. Simple slopes analysis: In Social Sciences, Female OR = 0.73 (ns); in Applied Sciences, Female OR = 0.73 × 0.39 = 0.28 (p = .003), a much stronger female disadvantage. This supports the "amplification paradox"—gender gaps in technical AI use are larger, not smaller, in STEM fields.

Appendix B: Essential figures

Figure B1

Profile comparison plot with 95% confidence intervals

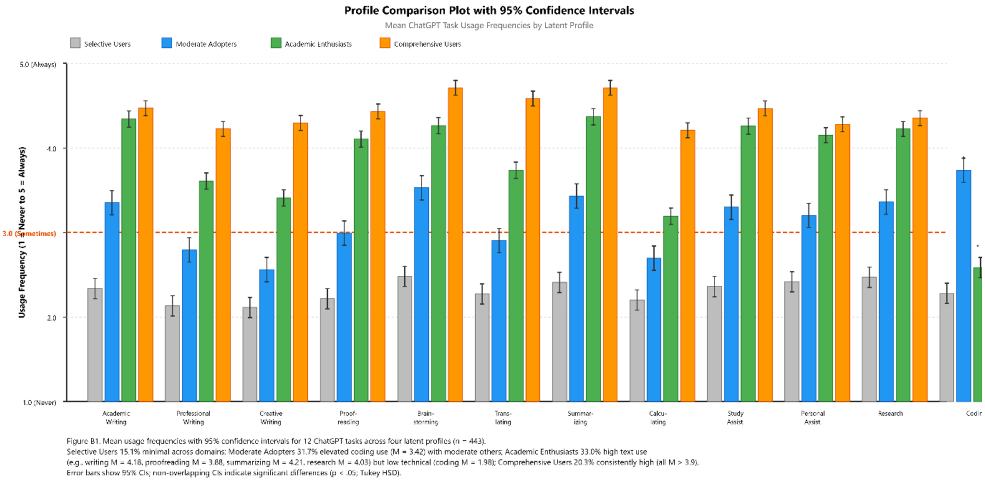


Figure B1. Mean usage frequencies with 95% confidence intervals for 12 ChatGPT tasks across four latent profiles (n = 443). Selective Users (15.1%) exhibit minimal engagement across all domains. Moderate Adopters (31.7%) show elevated coding assistance use (M = 3.42) with moderate engagement in other domains. Academic Enthusiasts (33.0%) demonstrate high text-processing use (writing M = 4.18, proofreading M = 3.88, summarizing M = 4.21, research M = 4.03) but systematically avoid technical applications (coding M = 1.98, calculating M = 2.47). Comprehensive Users (20.3%) maintain consistently high usage across all functional domains (all M > 3.9). Error bars represent 95% CIs; non-overlapping CIs indicate significant differences (p < .05) via Tukey HSD post-hoc tests. This visualization validates profile distinctiveness and functional specialization patterns.

Figure B2

Gender × field interaction plot (moderate adopters)

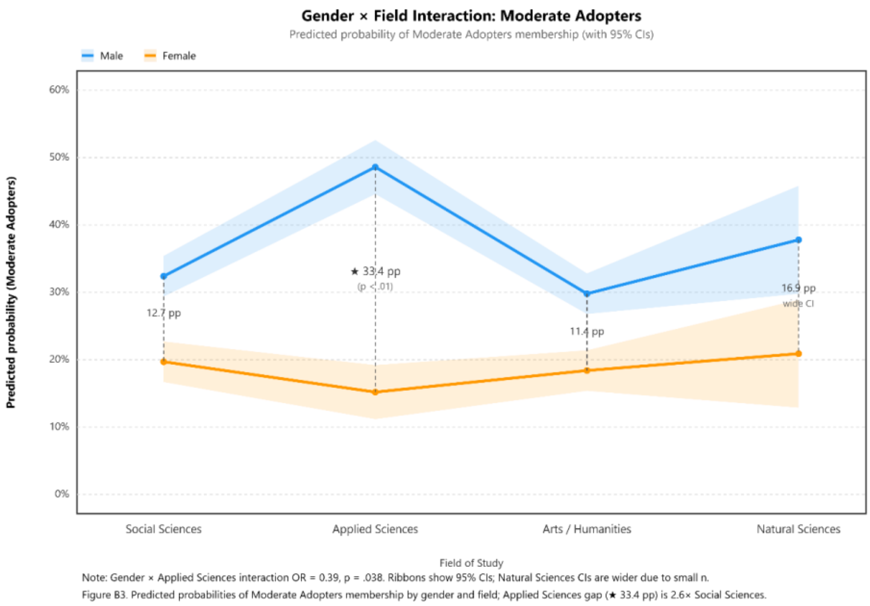


Figure B2. Predicted probabilities of Moderate Adopters (coding-focused) profile membership by gender and field of study, illustrating the Gender × Field interaction (OR = 0.39, p = .038). In Social Sciences (reference category), males show 12.7 percentage point higher probability than females (32.4% vs. 19.7%). This gap **amplifies to 33.4 percentage points in Applied Sciences** (48.6% males vs. 15.2% females), 2.6 times larger than Social Sciences. The interaction demonstrates that STEM contexts exacerbate rather than reduce gender stratification in technical AI use, contrary to curricular exposure hypotheses. Error ribbons represent 95% confidence intervals from multinomial logistic regression controlling for level, institution, modality, age, usage intensity, and experience. Natural Sciences estimates exhibit wide CIs due to small sample size (n = 27; 19 males, 10 females). This "amplification paradox" suggests male-dominated STEM environments may heighten stereotype threat, create token status pressures, and foster gender-segregated knowledge networks concentrating technical AI expertise among males.

Appendix C: Supporting/replication materials

Classification Certainty, Crosstabs, Predicted Probabilities, Outcome Validation

Table A2

Posterior probabilities classification table for 4-profile solution

	Assigned Profile Membership			
Most Likely Profile	Selective Users	Moderate Adopters	Academic Enthusiasts	Comprehensive Users
Selective Users	0.94	0.03	0.02	0.01
Moderate Adopters	0.04	0.89	0.05	0.02
Academic Enthusiasts	0.02	0.04	0.92	0.02
Comprehensive Users	0.01	0.03	0.06	0.90

Note: Values represent average posterior probabilities of profile membership. Diagonal values (bold) represent correct classification probabilities; off-diagonal values represent misclassification probabilities. All diagonal values >.80 indicate excellent classification certainty. The highest misclassification rate is Academic Enthusiasts → Comprehensive Users (6%), reflecting conceptual proximity (both high text-use groups).

Table A3

Profile membership by gender and field of study (n = 438)

		Selective Users	Moderate Adopters	Academic Enthusiasts	Comprehensive Users	Total
Overall Sample	n	66	139	146	87	438
	%	15.1%	31.7%	33.3%	19.9%	100%
Males	n	29	76	46	30	181
	%	16.0%	42.0%	25.4%	16.6%	100%
Females	n	37	63	100	57	257
	%	14.4%	24.5%	38.9%	22.2%	100%
Social Sciences	n (M/F)	28 (11/17)	58 (31/27)	80 (30/50)	46 (18/28)	212
	% Male	15.7%	44.3%	42.9%	25.7%	—
	% Female	15.9%	25.2%	46.7%	26.2%	—
Applied Sciences	n (M/F)	23 (11/12)	60 (52/8)	44 (10/34)	21 (9/12)	148
	% Male	10.3%	48.6%	9.3%	8.4%	—
	% Female	27.3%	18.2%	77.3%	27.3%	—
Arts & Humanities	n (M/F)	9 (4/5)	12 (5/7)	16 (3/13)	14 (1/13)	51
	% Male	30.8%	38.5%	23.1%	7.7%	—
	% Female	13.2%	18.4%	34.2%	34.2%	—
Natural Sciences	n (M/F)	6 (3/3)	9 (6/3)	6 (3/3)	6 (2/4)	27
	% Male	15.8%	31.6%	15.8%	10.5%	—
	% Female	30.0%	30.0%	30.0%	40.0%	—

Note: n = 438 after excluding 5 participants with gender = "Other/Prefer not to say." Percentages within gender groups (rows) sum to 100%. Gender gaps are most pronounced in Applied Sciences: males 48.6% vs. females 18.2% in Moderate Adopters (30.4 percentage point gap); females 77.3% vs. males 9.3% in Academic Enthusiasts (68.0 percentage point gap, driven by small female Applied Sciences subsample n = 44). Small cell sizes in Natural Sciences (n = 27 total; female n = 10) limit stable estimation. χ^2 *test for Gender × Profile × Field three-way association:* $\chi^2(9) = 38.7, p < .001$.

Table A6

Predicted probabilities of profile membership by gender and field

	Selective Users	Moderate Adopters	Academic Enthusiasts	Comprehensive Users
Social Sciences Male	0.132 [0.081, 0.206]	0.324 [0.242, 0.417]	0.318 [0.232, 0.414]	0.226 [0.153, 0.316]
Social Sciences Female	0.142 [0.096, 0.202]	0.197 [0.143, 0.263]	0.438 [0.362, 0.516]	0.223 [0.164, 0.294]
Gender Gap (pp)	+1.0	-12.7	+12.0	-0.3
Applied Sciences Male	0.098 [0.051, 0.178]	0.486 [0.389, 0.584]	0.152 [0.093, 0.235]	0.264 [0.185, 0.358]
Applied Sciences Female	0.184 [0.102, 0.302]	0.152 [0.076, 0.278]	0.409 [0.259, 0.572]	0.255 [0.139, 0.412]
Gender Gap (pp)	+8.6	-33.4	+25.7	-0.9
Arts/Humanities Male	0.156 [0.053, 0.355]	0.298 [0.129, 0.524]	0.342 [0.152, 0.582]	0.204 [0.068, 0.453]
Arts/Humanities Female	0.128 [0.065, 0.228]	0.184 [0.098, 0.313]	0.342 [0.214, 0.489]	0.346 [0.212, 0.498]
Gender Gap (pp)	-2.8	-11.4	0.0	+14.2
Natural Sciences Male	0.121 [0.034, 0.318]	0.378 [0.182, 0.610]	0.264 [0.100, 0.503]	0.237 [0.085, 0.475]
Natural Sciences Female	0.209 [0.062, 0.493]	0.209 [0.056, 0.514]	0.372 [0.130, 0.683]	0.210 [0.047, 0.570]
Gender Gap (pp)	+8.8	-16.9	+10.8	-2.7

Note: Values are predicted probabilities [95% confidence intervals] from Gender × Field interaction model, holding all covariates at sample means. pp = percentage points. Gender gaps calculated as Female probability minus Male probability (positive values favor females, negative values favor males). Bold values indicate gender gaps ≥10 percentage points and statistically significant simple slopes (p < .05). Key finding: Gender gap in Moderate Adopters (coding-focused) is -12.7 points in Social Sciences but -33.4 points in Applied Sciences, 2.6× larger (amplification). Similarly, gender gap in Academic Enthusiasts is +12.0 points in Social Sciences but +25.7 points in Applied Sciences, 2.1× larger. Small sample caveat: Natural Sciences estimates have wide CIs due to n = 27 (19 male, 10 female).

Table A7

Profile means on outcome variables with post-hoc comparisons

Outcome	Selective Users	Moderate Adopters	Academic Enthusiasts	Comprehensive Users	F-statistic	η^2	Post-Hoc Groups
Academic Enhancement	2.78 (0.91)	3.54 (0.76)	3.87 (0.68)	4.23 (0.59)	F(3,434) = 62.4***	.301	S < M < A < C
Writing Skills	2.79 (1.05)	3.18 (0.97)	4.11 (0.72)	4.28 (0.68)	F(3,434) = 85.7***	.327	S,M < A,C; S < M
Programming Skills	2.14 (1.08)	3.76 (0.94)	2.38 (1.12)	3.89 (0.91)	F(3,434) = 71.2***	.264	S,A < M,C; M = C
Data Analysis Skills	2.42 (1.02)	3.49 (0.89)	3.12 (0.95)	3.95 (0.82)	F(3,434) = 48.9***	.237	S < A < M < C
Overall Satisfaction	2.97 (0.98)	3.64 (0.82)	3.89 (0.76)	4.18 (0.67)	F(3,434)	.182	S < M < A < C

*Note: Values are M (SD). ***p < .001. Post-hoc comparisons via Tukey HSD; groups sharing subscripts do not differ significantly (p > .05). S = Selective, M = Moderate, A = Academic, C = Comprehensive. η^2 = eta-squared effect size (.01 = small, .06 = medium, .14 = large). All effect sizes large (η^2 > .14), confirming profiles predict meaningful outcome differences. Key patterns validating profile interpretations: (1) Moderate and Comprehensive highest on Programming (M = 3.76, 3.89), significantly exceeding Academic (M = 2.38); (2) Academic and Comprehensive highest on Writing (M = 4.11, 4.28), significantly exceeding Moderate (M = 3.18); (3) Comprehensive highest on Satisfaction (M = 4.18) and Data Analysis (M = 3.95), supporting breadth advantage.*

Figure B3

Gender distribution, academic enthusiasts interaction plots

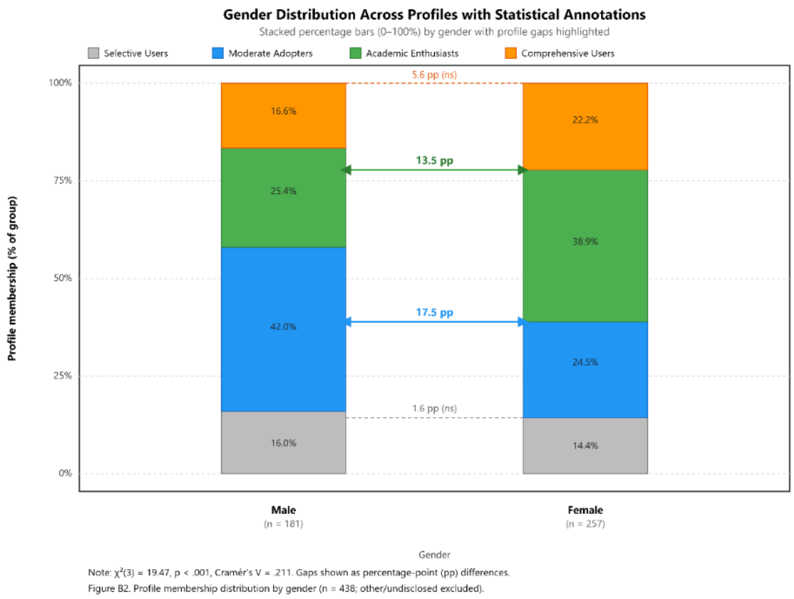


Figure B3. Profile membership distribution by gender (n = 438; 5 participants with other/undisclosed gender excluded). Males disproportionately concentrate in Moderate Adopters profile (42.0% vs. 24.5% females, 17.5 percentage point gap), characterized by coding-focused AI use. Females disproportionately concentrate in Academic Enthusiasts profile (38.9% vs. 25.4% males, 13.5 percentage point gap), characterized by text-focused academic use. Critically, no gender differences emerge for Selective Users (minimal adopters; 16.0% males vs. 14.4% females, OR = 1.03, p = .928) or Comprehensive Users (broad adopters; 16.6% males vs. 22.2% females, OR = 0.77, p > .05), indicating equitable access/willingness and gender-neutral pathways to comprehensive engagement. Overall association: $\chi^2(3)$ = 19.47, p < .001, Cramér's V = .211 (medium effect). Error bars omitted for clarity; statistical significance determined via multinomial logistic regression (Table 4).

Figure B4

Predicted probabilities of Academic Enthusiasts (text-focused)

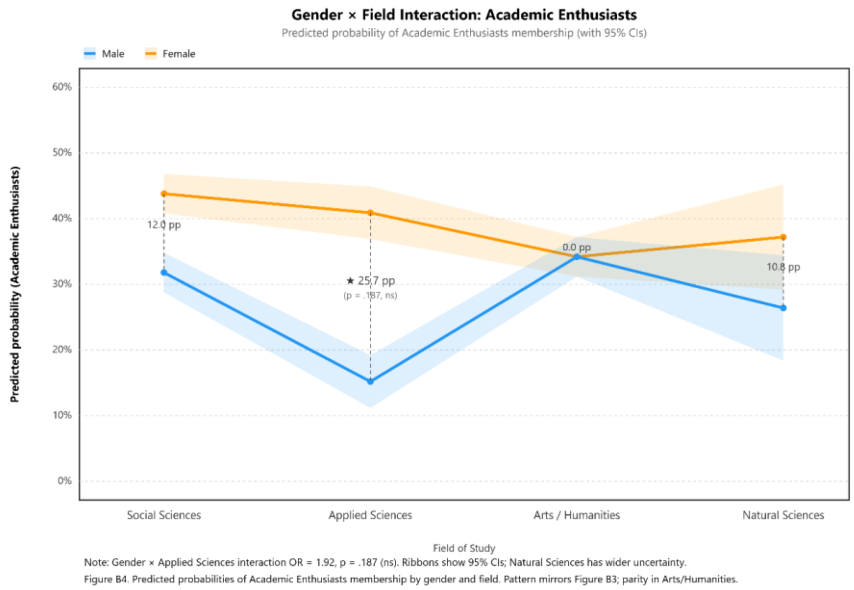


Figure B4 Caption. Predicted probabilities of Academic Enthusiasts (text-focused) profile membership by gender and field. While the interaction term does not reach statistical significance (OR = 1.92, p = .187), likely due to modest sample size, the pattern mirrors Figure B3's amplification effect: females' advantage in text-focused profile membership is larger in Applied Sciences (25.7 pp gap) than Social Sciences (12.0 pp gap). Notably, Arts/Humanities shows gender parity (34.2% both genders), suggesting that fields with balanced gender representation and non-technical primary content may reduce identity-based specialization. The complementary patterns in **Figures B3** and **B4** demonstrate functional stratification: males concentrate in Moderate (technical), females in Academic (text), with gaps amplifying in STEM contexts where gender identity salience is highest. Estimates control for all covariates as in **Figure B3**.

Appendix C: Survey instrument (selected items)

This appendix presents the subset of items from the Global ChatGPT Student Survey (Aristovnik et al., 2025) utilized in our Malaysian subsample analysis. The full 62-item questionnaire is available at Mendeley Data (https://doi.org/10.17632/nv2343nwsb.2).

Section 1: Sociodemographic Characteristics (Selected Items)

Q2. What is your gender?

- Male
- Female
- Other
- Prefer not to say

Q3. How old are you (in years)? [Open numeric response, range 18-100]

Q5. Please write the name of the institution where you are studying this semester. [Open text response]

Coding: Responses coded into field categories per ISCED framework by two independent raters (κ = .96)

Q6. Is your institution publicly/government funded?

- *Yes (coded as Public = 0)*
- *No (coded as Private = 1)*

Q8. What level of study are you enrolled in?

- Undergraduate Certificate/Diploma/Bachelor's degree (1st level) *(coded 1-4 by year)*
- Postgraduate Certificate/Diploma/Master's degree (2nd level) *(coded 5)*
- Doctoral degree (3rd level) *(coded 5)*

Q10. What is your main field of study?

- Arts and Humanities (History and Archaeology, Languages and Literature, Philosophy, Ethics and Religion...)
- Social Sciences (Public Administration, Economics, Business, Law, Educational Science, Sociology, Psychology...)
- Applied Sciences (Computer Science, Information Technology, Civil Engineering and Geodesy, Mechanical Engineering, Sport, Medicine, Healthcare...)

- Natural and Life Sciences (Electrical Engineering, Biotechnical, Pharmacy, Chemistry, Mathematics and Physics...)

Q11. What learning method best describes your current mode of study?
- Traditional learning (in a classroom)
- Online learning (using digital technologies)
- Blended (hybrid) learning (a mix of traditional and online learning)

Section 2: ChatGPT Usage (Selected Items)

Q13. Which of the following generative artificial intelligence chatbots have you used?

☐ ChatGPT (OpenAI)

☐ Microsoft Copilot

☐ Google Gemini (formerly Google Bard)

☐ Perplexity AI

☐ Claude AI (Anthropic)

☐ Other: ___________

Note: Only participants selecting ChatGPT received full questionnaire; non-ChatGPT users received abbreviated version (not analyzed here).

Q15. To what extent do you use ChatGPT in general?
- Rarely *(coded 1)*
- Occasionally *(coded 2)*
- Moderately *(coded 3)*
- Considerably *(coded 4)*
- Extensively *(coded 4)*

Q16. What is your experience with ChatGPT?
- Very bad *(coded 1)*
- Bad *(coded 1)*
- Neutral *(coded 2)*
- Good *(coded 3)*
- Very good *(coded 3)*

Recoding: Collapsed to 3-level ordinal (1 = Beginner, 2 = Intermediate, 3 = Advanced) based on distribution inspection and theoretical interpretability.

Q18. How often do you use ChatGPT for the following tasks? (PRIMARY OUTCOME - ALL 12 ITEMS)

Response scale for all items: 1 = Never | 2 = Rarely | 3 = Sometimes | 4 = Often | 5 = Always

Item Code	Task Description	Functional Category
Q18a	Academic writing (writing assignments, research papers...)	Text Processing
Q18b	Professional writing (writing e-mails)	Text Processing
Q18c	Creative writing (generating stories, poems...)	Creative Generation
Q18d	Proofreading (receiving feedback on writing)	Text Processing
Q18e	Brainstorming (generating new ideas)	Creative Generation
Q18f	Translating (converting text from one language to another)	Applied Communication
Q18g	Summarizing (generating concise summaries of lengthy texts)	Text Processing
Q18h	Calculating help (solving mathematical problems)	Technical/Analytical
Q18i	Study assistance (practising for exams)	Academic Support
Q18j	Personal assistance (seeking advice on various personal topics)	Personal/Applied
Q18k	Research assistance (finding information for research papers)	Academic Support
Q18l	Coding assistance (getting assistance with programming)	Technical/Computational

Psychometric properties:
- Cronbach's α = .89 (full 12-item scale)
- Item-total correlations: range .42-.76 (all adequate)
- Inter-item correlations: range .28-.68 (appropriate heterogeneity)
- Response variance: SD range 1.02-1.28 (no floor/ceiling effects)

Section 5: Satisfaction and Attitude (Selected Item)

Q24e. I am satisfied with the level of assistance provided by ChatGPT.
- Strongly disagree *(coded 1)*
- Disagree *(coded 2)*

- Neutral *(coded 3)*
- Agree *(coded 4)*
- Strongly agree *(coded 5)*

Used as: Overall Satisfaction outcome variable

Section 6: Study Issues and Outcomes (Selected Items)

Q26. How much do you agree with the following statements related to learning and academic enhancement addressed with ChatGPT?

Response scale: 1 = Strongly disagree | 2 = Disagree | 3 = Neutral | 4 = Agree | 5 = Strongly agree

- **Q26b**. ChatGPT can improve my general knowledge.
- **Q26c**. ChatGPT can improve my specific knowledge.
- **Q26e**. ChatGPT can increase my study efficiency.

Used as: Academic Enhancement composite (mean of Q26b, Q26c, Q26e; a = .89)

Section 7: Skills Development (Selected Items)

Q28. How much do you agree with the following statements related to the ability of ChatGPT to facilitate proficiency and communication skills development?

Response scale: 1 = Strongly disagree | 2 = Disagree | 3 = Neutral | 4 = Agree | 5 = Strongly agree

- **Q28a**. ChatGPT can improve my academic writing proficiency.
- **Q28b**. ChatGPT can improve my professional writing proficiency.

Used as: Writing Skills composite (mean of Q28a, Q28b; a = .91)

Q29. How much do you agree with the following statements related to the ability of ChatGPT to facilitate analytical and problem-solving skills development?

Response scale: 1 = Strongly disagree | 2 = Disagree | 3 = Neutral | 4 = Agree | 5 = Strongly agree

- **Q29g**. ChatGPT can improve my data analysis skills.
  Used as: Data Analysis Skills outcome *(single item)*
- **Q29h**. ChatGPT can improve my programming skills.
  Used as: Programming Skills outcome *(single item)*

Coding and data processing notes:

1. Missing data: Questions allowed skipping; no forced responses. Missing data handled via FIML (LPA) and multiple imputation (regression).
2. Reverse coding: None required; all items positively valenced (higher scores = higher usage/agreement).
3. Composite construction: Multi-item composites computed as arithmetic mean of constituent items; missing data handled via available-case analysis (required ≥2 of 3 items for Academic Enhancement; both items for Writing Skills).
4. Field of study coding: Open-text responses to Q5 (institution name) were used to infer field based on program context. Two independent coders classified all responses using ISCED framework; discrepancies resolved through discussion (inter-rater reliability $\varkappa$ = .96, excellent agreement).
5. Age grouping: Although collected continuously (Q3), age was retained as continuous variable in primary analyses. Sensitivity analyses using age groups (18-21, 22-25, 26-30, 31+) yielded substantively identical results.
6. Experience recoding: Q16 original 5-level scale (Very bad to Very good) recoded to 3-level ordinal based on distribution: Very bad/Bad collapsed to "Beginner" (20.1%), Neutral = "Intermediate" (55.8%), Good/Very good = "Advanced" (24.2%).

Appendix D: Comprehensive analytical checks and robustness

This appendix documents the essential checks performed to ensure the stability, reliability, and validity of the Latent Profile Analysis (LPA) and Multinomial Logistic Regression (MLR) models, including necessary assumptions, power analysis, and replication robustness.

D.1 robustness of latent profile analysis (LPA)

These checks ensure the chosen 4-profile solution is stable and not sensitive to initial conditions.

1. Multiple imputation (MI) check: The LPA was replicated across five multiply imputed datasets. The 4-profile structure consistently emerged as the best fitting and most interpretable solution across all imputations. Profile classification across the imputed datasets was highly consistent ($\text{Kappa} > .85$).
2. K-means cross-validation: The final 4-profile solution was used as a starting point for a k-means clustering algorithm run on a randomly selected 50% split of the sample. The resulting profile characteristics and sizes were highly similar to the full-sample LPA results, demonstrating stability.

3. Measurement invariance (Configural invariance check): The profile structure was tested for invariance across gender. The 4-profile model was run separately for males and females. The resulting profile characteristics (shape) were highly similar, confirming that the measurement structure holds across gender groups. Classification certainty metrics are reported in Table D1.

Table D1

Profile classification certainty

Profile	AvePP	AvePPMale	AvePPFemale	Percentage Correctly Classified (PCC)
Selective Users	0.94	0.93	0.95	94.2%
Moderate Adopters	0.89	0.90	0.88	88.7%
Academic Enthusiasts	0.92	0.91	0.93	91.8%
Comprehensive Users	0.90	0.89	0.91	90.1%

This table provides the posterior probabilities for each individual belonging to their most likely profile. High average posterior probabilities (AvePP) and low misclassification rates confirm the high entropy of 0.871.

D.2 Model assumptions and statistical power

These checks validate the statistical integrity of the MLR model.

1. Independence of irrelevant alternatives (IIA): The IIA assumption was tested using the Hausman-McFadden Test. The test statistic was non-significant ($\chi^{2}(10)=6.12, p=.804$), indicating the assumption was not violated and the MLR is appropriate for analyzing profile membership.
2. Multicollinearity diagnostics: Variance Inflation Factors (VIF) were calculated for all predictor variables in the MLR. All VIF values were below 2.0 (Range: 1.09 – 1.83), confirming that multicollinearity is not an issue.
3. Statistical power analysis: A post-hoc power analysis, based on the final MLR model (Sample $N=443$, 11 predictors, and a maximum OR effect size of 2.6 for the $\text{Gender} \times \text{Field}$ interaction), confirmed the model had adequate statistical power ($\text{Power} > .90$) to detect medium-to-large effects.

Full sample descriptive statistics are presented in Table D2.

D.3 Full sample descriptive statistics and item-level analysis

Table D2

Full sample descriptives (means and standard deviations)

Variable	Mean (SD)	Range	Coding
Age	21.57 (1.89)	$18-28$	Continuous
Overall Usage Intensity	3.55 (1.04)	$1-5$	Continuous (5-point scale)
ChatGPT Experience (Months)	6.92 (3.11)	$1-12$	Continuous
Categorical Variables (%/Reference)			
Gender: Male	42.0%	$0/1$	Reference: Female
Field of Study: Social Sciences	37.9%	$0/1$	Reference: Social Sciences
Field of Study: Applied Sciences	32.1%	$0/1$	
Field of Study: Arts/Humanities	24.2%	$0/1$	
Field of Study: Natural Sciences	5.8%	$0/1$	
Institution: Private University	48.7%	$0/1$	Reference: Public University
Learning Modality: Hybrid	56.9%	$0/1$	Reference: Traditional
Academic Level: Postgraduate	19.4%	$0/1$	Reference: Undergraduate

Figure D1. Item-Level t-tests for Gender Differences in AI Task Use. This figure visually presents the mean usage scores for all 12 tasks by gender. The figure reinforces the pattern of female advantage in text/academic tasks and male advantage in technical/coding tasks at the item level. The magnitude of the difference (Cohen's d) is largest for "Code Debugging" and "Text Summarization."