



## OPEN Associations of task technology fit and perceived usefulness with responsible generative AI use among university students

Qiufen Wang<sup>1</sup>, Ying Wang<sup>2</sup>✉ & Mansour Amini<sup>1,3</sup>

This study examines how task–technology fit (TTF) and perceived usefulness (PU) are associated with responsible generative AI use (RGU) in higher education. Survey data from 280 university students in China were analyzed using covariance-based structural equation modeling. The model showed excellent fit, and all hypothesized paths were positive and significant: TTF → PU ( $\beta = 0.360$ ), TTF → RGU ( $\beta = 0.318$ ), and PU → RGU ( $\beta = 0.273$ ). Bootstrap results indicated a partial mediation pattern in which PU was associated with the relationship between TTF and RGU. Females reported higher RGU than males, doctoral students outscored undergraduates on TTF and PU, and frequent GenAI users scored highest across all constructs. The findings extend TTF and TAM by integrating responsibility as a behavioral outcome and indicate that task-aligned, transparent GenAI practices may support more sustainable learning.

**Keywords** Generative AI, Task–Technology fit, Perceived usefulness, Responsible GenAI use, Higher education

The rapid development of generative artificial intelligence (GenAI) tools such as ChatGPT and other large language model based systems is reshaping higher education. Students increasingly use these tools to support academic tasks such as information retrieval, text summarisation, essay drafting, and data analysis, and recent surveys point to widespread uptake for writing and research support<sup>13,31</sup>. Alongside these opportunities, universities are grappling with concerns about academic integrity, ethical use, and sustainable adoption<sup>27</sup>. Many institutions have started issuing guidelines for responsible GenAI use<sup>34</sup>, yet students' actual practices, disclosure habits, and perceptions of responsibility remain uneven.

Technology-adoption research has begun to examine GenAI through frameworks such as the Technology Acceptance Model and the Unified Theory of Acceptance and Use of Technology, highlighting the importance of perceived usefulness, performance expectancy, and facilitating conditions for students' willingness to adopt AI tools<sup>5,30</sup>. At the same time, empirical and conceptual work points to ethical risks related to privacy, bias, plagiarism, and unclear policies in educational contexts<sup>24,26,32</sup>. Recent discussions therefore argue that Responsible GenAI Use (RGU)—including disclosure of AI assistance, adherence to institutional policies, and integrity-consistent behaviour—should be treated as a key outcome of GenAI integration<sup>2,8</sup>. However, adoption models such as TAM and UTAUT were designed primarily to predict intentions and actual usage, and they say relatively little about the quality or responsibility of that use. This creates a gap between adoption-focused frameworks and emerging work on ethical GenAI practices, especially around disclosure and academic integrity.

To address this gap, the present study develops and tests a structural equation model that links task–technology fit (TTF), perceived usefulness (PU), and responsible GenAI use (RGU) among higher-education students. TTF is defined here as the perceived alignment between GenAI functionalities and academic task requirements. Grounded in Task–Technology Fit theory<sup>16</sup> and the Technology Acceptance Model<sup>12</sup>, the model examines whether students who perceive better task–technology alignment also view GenAI as more useful, and whether these perceptions are associated with more responsible practices such as disclosure, policy alignment, and verification. In addition, subgroup differences by gender, programme level, and frequency of GenAI use are explored to provide contextual evidence on how responsibility patterns vary across student groups. The hypotheses derived from this framework are presented after the literature review and conceptual model.

<sup>1</sup>School of Foreign Languages, Kashi University, Kashi 844000, China. <sup>2</sup>School of Humanities, Wuhan University of Engineering Science, Jiangxia District, Wuhan 430200, Hubei, China. <sup>3</sup> School of Languages, Literacies and Translation, Universiti Sains Malaysia, Penang 11800, Malaysia. ✉email: judithwy37@163.com

This study contributes to the emerging literature on GenAI in higher education in three ways. First, it extends TTF and TAM by integrating them with a behavioural construct of responsible GenAI use, thereby embedding ethical engagement directly into an established adoption framework. Second, it provides empirical evidence from a multi-level student sample (undergraduate, master's, and doctoral) using a validated measurement instrument and covariance-based structural equation modelling. Third, it offers practical implications for course and policy design, suggesting how universities can align GenAI tools with academic tasks while promoting disclosure, integrity, and sustainable learning practices.

Accordingly, the study is guided by the following research questions:

- 1) RQ1: How does task–technology fit (TTF) relate to students' perceived usefulness (PU) of GenAI tools in academic contexts?
- 2) RQ2: How do task–technology fit (TTF) and perceived usefulness (PU) associate with responsible GenAI use (RGU)?
- 3) RQ3: Does perceived usefulness (PU) mediate the association between task–technology fit (TTF) and responsible GenAI use (RGU)?
- 4) RQ4: How do students' perceptions of TTF, PU, and RGU vary across demographic variables such as gender, programme level, and frequency of GenAI use?

## Literature review

### Task–Technology fit in higher education

The Task–Technology Fit theory, introduced by Goodhue and Thompson<sup>16</sup>, asserts that technology is positively associated with individual performance when its functionalities align with task requirements. Within higher education, TTF has become an influential framework for analyzing the integration of emerging technologies, particularly generative artificial intelligence, into learning environments. In this tradition, TTF is typically operationalised through users' judgements of how well a system's features, information, and interface support the efficient completion of their core tasks.

Empirical evidence suggests that GenAI tools provide strong functional support for academic activities, yet their educational benefits remain constrained by gaps in task alignment and responsible use. Al Zaidy<sup>2</sup> reported that while 86% of students across 16 countries actively employ tools such as ChatGPT, Grammarly, and Microsoft Copilot, only 5% are aware of institutional AI policies, and more than half express concerns regarding fairness, privacy, and over-reliance. This indicates that although GenAI aligns with key academic tasks such as writing, summarization, and information retrieval, the lack of ethical literacy and institutional guidance reduces its effectiveness, underscoring the relevance of TTF as a lens for assessing sustainable integration. In the present study, however, these ethical, fairness, and policy concerns are not treated as dimensions of TTF itself but as part of the broader context that motivates the separate construct of Responsible GenAI Use.

Beyond GenAI, reviews of technology adoption reinforce TTF's centrality in educational research. Jebril et al.<sup>22</sup> found that TTF has been extensively applied in digital libraries, mobile learning, and educational management systems, consistently showing that strong alignment between task demands and technological features enhances user satisfaction, utilization, and performance. Compared with the Information System Success Model (ISSM), which emphasizes service quality and user satisfaction, TTF offers a more direct measure of the practical fit between technology functions and academic needs. However, integrated applications of ISSM and TTF remain limited, suggesting the need for more holistic frameworks that capture both system-level quality and task alignment.

Recent empirical studies provide further support for TTF's predictive power in GenAI contexts. Al-Mamary et al.<sup>6</sup>, using survey data from Saudi universities, demonstrated that task, technology, and individual characteristics jointly shape students' perceived fit with ChatGPT, which in turn strongly predicts behavioural intentions to adopt it in learning. Similarly, Alyoussef<sup>7</sup> extended TTF by integrating it with the Technology Acceptance Model (TAM), showing that perceived usefulness, ease of use, and enjoyment significantly contribute to stronger task–technology alignment, thereby enhancing e-learning acceptance and student satisfaction. Al-Dokhny et al.<sup>4</sup> further revealed that multimodal large language models (MLLMs) positively affect performance and adoption when task and technology features are well aligned, although social influence and hedonic motivation may weaken effort expectancy.

Taken together, these studies confirm that TTF provides a robust framework for understanding technology adoption in higher education. They also highlight a crucial gap: while TTF research has largely focused on performance, satisfaction, and adoption, limited attention has been paid to its connection with responsible GenAI use. Positioning TTF within this ethical dimension extends its explanatory scope, linking technological affordances not only to efficiency and adoption but also to sustainable and responsible learning practices. Consistent with this focus, the present study operationalises TTF through items TTF1–TTF5 that ask students to evaluate how effectively GenAI tools support their academic tasks in terms of functional adequacy, efficiency, and overall task alignment, while ethical literacy, fairness, and institutional guidance are captured through the distinct RGU construct rather than through the TTF scale.

### Perceived usefulness and technology adoption

Perceived Usefulness, one of the core constructs of the Technology Acceptance Model developed by Davis<sup>12</sup>, refers to the extent to which users believe that technology enhances their performance and efficiency. Across diverse contexts, PU has consistently emerged as a decisive determinant of technology adoption, shaping not only behavioural intention (BI) but also long-term integration into professional and educational practices.

In school settings, Oktaria et al.<sup>29</sup> demonstrated that both PU and perceived ease of use (PEOU) significantly influenced attitudes toward technology (ATU), which in turn predicted BI. Interestingly, teachers reported

stronger perceptions of usefulness than students, indicating that educators see greater pedagogical value in digital tools. This finding suggests that enhancing PU among students is crucial for sustainable adoption of educational technologies, a point directly relevant to higher education institutions considering the integration of GenAI.

Similar conclusions are found beyond education. Alamir<sup>3</sup>, in a study of digital banking adoption in Saudi Arabia, confirmed that PU and PEOU strongly predicted both BI and actual usage, with PU exerting the strongest influence. Users who perceived digital banking as useful for improving efficiency and task performance were significantly more likely to adopt and sustain its use, whereas perceived risk acted as a deterrent. Translating these insights to higher education suggests that students are more likely to adopt GenAI responsibly if they recognize clear academic benefits, such as improved efficiency in research, writing, and problem-solving.

The mediating role of attitude has also been widely confirmed. Geddam et al.<sup>14</sup> showed that PU not only directly influenced BI but also indirectly strengthened adoption through positive user attitudes. In the context of AI tools, PU shaped both cognitive evaluations and affective dispositions, making it a pivotal driver of adoption. This dual pathway indicates that in higher education, strategies to promote GenAI adoption should highlight both functional benefits and positive affective experiences, thereby aligning usefulness with responsible engagement.

Systematic evidence also emphasizes PU's robustness. Aurangzeb et al.<sup>9</sup>, through a review of 30 studies, identified PU and PEOU as the most consistent predictors of Information and Communication Technology adoption among teachers, outweighing demographic factors such as age or gender. Their review further emphasized that institutional support and training significantly enhance perceptions of usefulness, thereby increasing adoption. Applied to GenAI in higher education, this suggests that beyond tool design, structured training and policy guidance are essential to strengthen students' perceptions of GenAI's usefulness for academic tasks.

Moreover, PU's significance extends even to non-traditional learners. Toros et al.<sup>33</sup>, studying older adults in lifelong learning programs, found that PU significantly shaped attitudes toward technology use, with PEOU mediating and gerontechnology self-efficacy moderating this relationship. This indicates that perceptions of usefulness remain critical across age groups, but their effect depends on ease of use and self-confidence in engaging with technology. In higher education, this highlights the importance of designing GenAI tools that are not only useful but also intuitive and supported by user training to build digital self-efficacy.

Taken together, these findings consolidate PU's central role in technology adoption. Whether in schools, banking, workplaces, or lifelong learning, PU consistently emerges as the strongest predictor of intention and usage. For higher education, this means that the responsible adoption of GenAI will depend heavily on students perceiving its tangible academic benefits, supported by institutional policies, training, and user-friendly design that collectively strengthen both cognitive and affective dimensions of acceptance.

### Responsible GenAI use in academic contexts

The notion of responsible use of generative artificial intelligence in higher education extends beyond technical adoption to encompass ethical, pedagogical, and institutional considerations. Gutw, Sorg, and Cruz Rodriguez<sup>17</sup> define responsible AI use as a comprehensive framework that embeds ethical principles and risk management across the entire AI lifecycle, emphasizing accuracy, inclusivity, transparency, and safety. Similarly, Wahle et al.<sup>35</sup> conceptualize RGU through three dimensions of transparency, integrity, and accountability, and propose the use of standardized "AI Usage Cards" to document disclosure, oversight, and responsibility in academic and scientific contexts. Taken together, these perspectives highlight that responsible GenAI use is not only about functional alignment with tasks but also about disclosure, human oversight, and institutional accountability to ensure sustainable and trustworthy practices.

Universities worldwide have begun to address these challenges through policy development. Dabis and Csáki<sup>11</sup> analyzed international policy documents and institutional guidelines, identifying four core ethical dimensions of accountability, human oversight, transparency, and inclusiveness that were reflected in university responses. Their study revealed a dual approach in which institutions emphasized individual responsibility and academic integrity, while at the same time allowing instructors autonomy in determining appropriate AI usage. Similarly, Atkinson-Toal and Guo<sup>8</sup> found substantial divergence among UK Russell Group universities, with some institutions adopting comprehensive frameworks to promote AI literacy and equitable access, whereas others provided only fragmented or minimal guidelines. These findings emphasize that responsible GenAI use in higher education requires standardized yet flexible policies that balance ethical safeguards with pedagogical innovation.

At the student level, research points to gaps between usage and awareness. Hou et al.<sup>20</sup> showed that while over 93% of computing students reported using ChatGPT at least once a month, many underestimated the frequency and intensity of their reliance. This discrepancy raises concerns about over-reliance, diminished peer interaction, and compromised academic integrity, highlighting the importance of self-awareness as a dimension of responsible use. Similarly, Liu, Park, and McMinn<sup>25</sup> found that while students appreciated GenAI's role in enhancing academic communication, they acknowledged limitations in supporting critical thinking and creativity, reinforcing the need for academic literacy interventions that go beyond tool usage.

Responsible GenAI use also intersects with issues of equity and digital literacy. Gesser-Edelsburg et al.<sup>15</sup> identified significant gaps in familiarity and confidence across gender, role, and discipline, showing that institutional and generational factors shape adoption. Chan and Lee<sup>10</sup> further demonstrated that Gen Z students were generally more optimistic about GenAI's potential, while Gen X and Gen Y teachers expressed greater concerns about ethical and pedagogical risks. These generational divides emphasize the importance of targeted training, critical thinking development, and evidence-based guidelines to bridge expectations and ensure equitable integration.

Taken together, the literature suggests that Responsible GenAI Use in higher education must be approached as a multidimensional construct that integrates individual responsibility, institutional policy, and ethical oversight. While universities are beginning to establish guidelines and students increasingly adopt these tools, gaps remain in policy consistency, student self-awareness, and equitable access. In this study, responsible GenAI use is therefore conceptualized at the student level as a pattern of ethical technology-use behaviour and academic integrity practice, reflected in tendencies to disclose AI assistance, comply with institutional guidelines, and critically verify AI-generated content. This view is compatible with broader frameworks of ethical technology use and AI literacy, which treat responsibility as a downstream behavioural manifestation of users' perceptions and competencies rather than as a purely normative prescription. Addressing these gaps is essential for aligning technological innovation with academic integrity, thereby paving the way for the conceptual framework outlined in the next section.

### Research gaps and conceptual framework

The literature review emphasizes the importance of task–technology fit, perceived usefulness, and responsible GenAI use in shaping students' engagement with emerging technologies. Prior studies show that TTF enhances adoption and performance outcomes, but they have generally focused on efficiency, satisfaction, or behavioural intention rather than responsible academic practices. Similarly, PU has been consistently validated as a decisive predictor of technology adoption (e.g., Alamir<sup>3,14</sup>), yet it remains unclear whether perceiving GenAI as useful also promotes ethical compliance, disclosure, and sustainable engagement.

At the same time, discussions of responsible GenAI use have largely taken place in institutional policy and ethics debates (e.g., Atkinson-Toal & Guo<sup>8,35</sup>), rather than in empirically tested behavioural models. Responsible use is often treated as a normative guideline, not as an outcome shaped by students' perceptions and experiences. As a result, adoption theories such as TAM and TTF remain weakly connected to ethical considerations that are now central to GenAI integration in higher education.

Building on these discussions, the present study conceptualizes responsible GenAI use as a student-level behavioural outcome that links ethical technology use, academic integrity, and AI literacy. The RGU construct captures intentions and self-reported tendencies to disclose AI assistance, comply with institutional guidelines, and critically evaluate AI-generated outputs, extending recent work on responsible GenAI behaviours in academic settings<sup>1</sup> and institutional AI policies<sup>8,35</sup>. This conceptualisation permits responsibility to be modelled in parallel with intention, adoption, and performance outcomes, while allowing TTF and PU to exert direct and indirect influences on responsibility within the same nomological network.

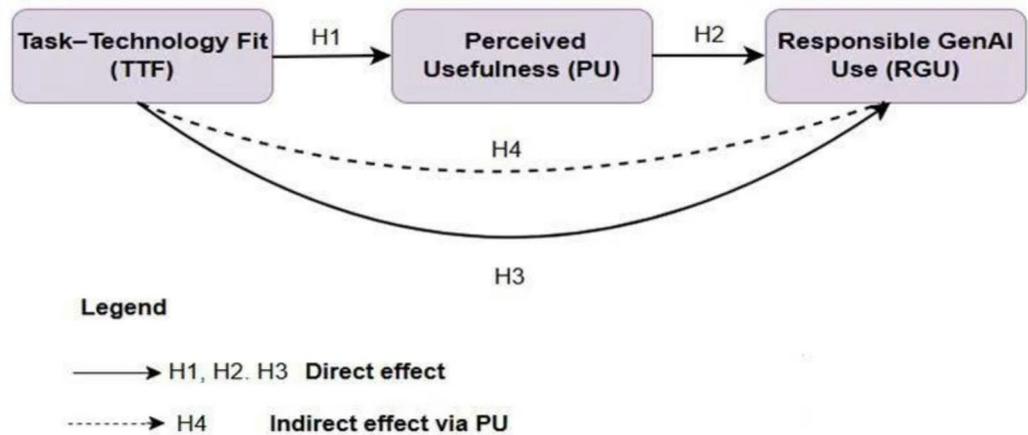
To address these gaps, this study develops a conceptual framework linking TTF, PU, and RGU. The framework is grounded in the Technology Acceptance Model<sup>12</sup> and the Task–Technology Fit theory<sup>16</sup>, while also drawing on recent studies that emphasize responsible and ethical use of GenAI in higher education (e.g., Al Maharmah et al.<sup>1</sup>). The model assumes that when students perceive that the functionalities of GenAI tools match the requirements of their academic tasks, they are more likely to find these tools useful. Students who perceive GenAI tools as useful are also expected to engage in more responsible practices, such as disclosing AI assistance, adhering to institutional policies, and complying with ethical standards. Within this framework, RGU is conceptualized as a behavioural intention construct adjacent to traditional TAM outcomes (e.g., intention and actual use), but with a specific focus on integrity-consistent engagement, disclosure, and policy alignment. This positioning allows ethical responsibility to be embedded directly within established adoption models, rather than treated as an external or purely normative consideration. In addition, TTF is expected to be directly associated with RGU, reflecting the idea that strong task–technology alignment may itself promote responsible engagement. When students experience a high degree of fit between GenAI functionalities and task requirements, the affordances and limits of the tools become more transparent, which can reduce ethical ambiguity about what counts as appropriate use. Clearer mapping between tools and pedagogical purposes is therefore expected to support boundary setting, disclosure, and verification as integral parts of responsible use. PU is further proposed as a mediator between TTF and RGU.

Beyond these structural relationships, demographic and behavioral differences may also play a role. Prior studies suggest that factors such as gender, academic level, and frequency of GenAI usage influence perceptions and practices of technology adoption (e.g., Chan & Lee<sup>10,15,20</sup>). On this basis, the present study examines group differences as theoretically informed, exploratory extensions of the core model, focusing on gender, programme level, and usage frequency as potential sources of systematic variation in TTF, PU, and RGU, without specifying separate group-level hypotheses.

Figure 1 presents the conceptual framework, which integrates adoption models with ethical considerations. By empirically testing this framework through structural equation modelling (SEM), the study extends existing theories of technology adoption and contributes to the growing discourse on responsible and sustainable GenAI use in higher education. In line with the four research questions, H1–H4 specify the structural relations among TTF, PU, and RGU, whereas RQ4 is addressed through exploratory subgroup comparisons (gender, programme level, and usage frequency) rather than through additional formal hypotheses.

Based on Fig. 1, the following hypotheses are formulated:

- H1** Task–Technology Fit is positively associated with Perceived Usefulness.
- H2** Perceived Usefulness is positively associated with Responsible GenAI Use.
- H3** Task–Technology Fit is positively associated with Responsible GenAI Use.
- H4** Perceived Usefulness mediates the association between Task–Technology Fit and Responsible GenAI Use.



**Fig. 1.** Proposed conceptual framework.

## Methods

### Research design

This study adopted a quantitative, cross-sectional survey design to examine the relationships among task-technology fit, perceived usefulness, and responsible generative AI use in a naturalistic higher-education context. Such a design is appropriate for theory-testing research that seeks to estimate latent constructs and structural relations at a single point in time without experimentally manipulating instructional conditions. The conceptual framework was guided by the Technology Acceptance Model, with TTF serving as an external determinant of PU and RGU as a behavioral intention construct focusing on disclosure, policy alignment, and integrity-consistent engagement. In addition to the core structural relations specified in RQ1–RQ3, the design also allowed for exploratory subgroup comparisons for RQ4.

Structural equation modeling was employed to validate the measurement model and to test the hypothesized structural relationships. SEM was selected because it allows for simultaneous estimation of latent constructs, measurement error, and mediation effects, providing a rigorous basis for modelling the indirect pathway from TTF to RGU through PU. All analyses were conducted in RStudio using packages such as lavaan, semTools, and psych.

### Participants and sampling

Participants were undergraduate, postgraduate, and doctoral students enrolled at universities in China who had prior experience using generative AI tools (e.g., ChatGPT, Bing Chat) for academic purposes. This inclusion criterion was adopted because the focal constructs in this study, namely task-technology fit, perceived usefulness, and responsible GenAI use, presuppose some firsthand experience with GenAI-supported academic tasks; students with no prior exposure would be unable to provide meaningful ratings of fit, usefulness, or responsible practices. Participants were recruited through multiple channels. The survey link hosted on Wenjuanxing was shared via course related online groups, and personal networks of university instructors who were invited to disseminate the study to their students. In addition, an open call for volunteers was posted on social media platforms such as Instagram to reach a broader pool of potential respondents. A purposive sampling method was therefore applied to specifically target students who were familiar with GenAI use in academic contexts. During data screening, responses to the frequency-of-use item were examined and cases indicating that students did not use GenAI for academic purposes were excluded, so that the final analytic sample comprised current or emerging GenAI users.

Following Kline's<sup>23</sup> guidelines of 10–20 participants per estimated parameter in structural equation modeling, a minimum of  $N = 280$  was set as the target sample size to ensure sufficient statistical power and the robustness of model estimation. At the same time, this sampling strategy necessarily excludes students who do not use GenAI, for example those who refrain due to ethical concerns, fear of misuse, limited digital literacy, or low technology readiness, so the findings should be interpreted as describing patterns among current or emerging GenAI users rather than the entire student population.

### Data collection

Data for this study were collected through an online questionnaire administered via Wenjuanxing, a widely used Chinese online survey platform. Consistent with the sampling strategy described above, the survey link was circulated through course-related online groups, university instructors' personal networks, and public posts on social media platforms to invite voluntary participation from students who used GenAI for academic purposes.

The questionnaire consisted of two sections. The first section included four demographic items, covering gender, age, program level, and frequency of using generative AI tools for academic purposes. The second section contained the main measurement scales adapted from established instruments. Task-technology

fit was measured with five items adapted from Goodhue and Thompson's<sup>16</sup> task–technology fit instrument, contextualized for GenAI-supported academic tasks. Perceived usefulness was assessed with four items adapted from Davis's<sup>12</sup> Technology Acceptance Model, reflecting students' perceptions of the academic utility of GenAI tools. Responsible generative AI use was measured with six items from the “Ethical Use of AI” factor of the AI Academic Integrity Scale<sup>1</sup>, capturing students' intention to disclose AI use, follow institutional guidelines, and maintain academic integrity when using GenAI. The full wording of all items, together with their original sources and minor contextual adaptations, is provided in Appendix to document precisely which indicators were taken from which instruments. All scale items were rated on a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree).

Prior to the main data collection, a pilot study was conducted with 50 student participants to evaluate the clarity, reliability, and validity of the questionnaire. The pilot results confirmed that the measurement model exhibited satisfactory reliability, with Cronbach's  $\alpha$  and composite reliability (CR) for all three constructs exceeding the recommended threshold of 0.70. The confirmatory factor analysis (CFA) further indicated an acceptable model fit ( $\chi^2/df = 1.24$ , CFI = 0.95, TLI = 0.94, RMSEA = 0.07), supporting the overall adequacy of the instrument for the subsequent large-scale survey.

In line with the SAGER guidelines, gender was collected as a self-reported demographic variable with binary response options (“male”, “female”), reflecting institutional reporting practices in the study context. Gender was included as a grouping variable because prior research on technology adoption and AI literacy has documented gendered patterns in access, attitudes, and usage. Throughout the analyses, any observed gender differences are interpreted cautiously as socioculturally and educationally shaped patterns of experience and opportunity rather than as biologically determined differences.

### Data analysis

Data were analyzed using RStudio with packages including lavaan, semTools, and psych. Prior to hypothesis testing, data screening procedures were conducted to check for missing values, outliers, and normality. For the fifteen indicator variables (TTF1–TTF5, PU1–PU4, RGU1–RGU6), no missing responses were observed (0% missingness), so the SEM analyses were based on the full sample of 280 cases. Multivariate outliers were screened using Mahalanobis distance; inspection of the distance distribution did not reveal extreme cases warranting deletion, and all observations were retained. Skewness and kurtosis statistics were examined to assess normality. Inspection of these indices did not indicate extreme non-normality, and, given the sample size ( $N = 280$ ), maximum likelihood estimation was considered appropriate because it is generally robust to mild deviations from multivariate normality in large samples. Descriptive statistics, including means, standard deviations, and Pearson correlations, were then computed for all study variables.

To evaluate potential common method bias, Harman's single-factor test and confirmatory factor-analytic comparisons were conducted. In an exploratory factor analysis with one factor, the first factor accounted for less than 40% of the total variance, below the conventional threshold that would suggest serious common method bias. In addition, a one-factor CFA model in which all fifteen items loaded on a single latent factor was specified and compared with the theorized three-factor model. The detailed fit indices for these models are presented in the Results section, but the overall pattern indicates that the one-factor solution fits substantially worse than the three-factor model, suggesting that common method variance is unlikely to be a dominant artifact in the observed associations among TTF, PU, and RGU.

The reliability and validity of the constructs were assessed through confirmatory factor analysis (CFA). Internal consistency was evaluated using Cronbach's  $\alpha$  and composite reliability (CR), with values of 0.70 or higher interpreted as indicating acceptable reliability (e.g., Nunnally & Bernstein<sup>28</sup>). Convergent validity was established by ensuring factor loadings exceeded 0.60 and the average variance extracted (AVE) exceeded 0.50. These criteria follow common recommendations in the SEM literature (e.g., Hair et al.,<sup>18</sup>). Discriminant validity was confirmed through the Fornell–Larcker criterion and the heterotrait–monotrait ratio (HTMT) being below 0.85. An HTMT value below 0.85 is generally taken to indicate adequate discriminant validity<sup>19</sup>.

Model fit was evaluated using  $\chi^2/df$ , the Comparative Fit Index (CFI), the Tucker–Lewis Index (TLI), the Root Mean Square Error of Approximation (RMSEA), and the Standardized Root Mean Square Residual (SRMR). Consistent with widely used SEM guidelines (e.g., Hu & Bentler<sup>21,23</sup>), CFI and TLI values of 0.95 or above (with 0.90 as a minimum acceptable level), RMSEA values of 0.06 or below (with values up to 0.08 considered acceptable), SRMR values of 0.08 or below, and  $\chi^2/df$  ratios below 3 were taken to indicate good model fit. To support the exploratory subgroup comparisons for RQ4, multi-group confirmatory factor analyses were conducted to evaluate measurement invariance across gender, program level, and frequency of GenAI use. Configural, metric, and scalar models were estimated using maximum likelihood with FIML and  $std.lv = TRUE$ . Invariance decisions followed conventional change-in-fit criteria ( $\Delta CFI \leq 0.010$ ,  $\Delta RMSEA \leq 0.015$ , and  $\Delta SRMR \leq 0.030$  for metric and  $\leq 0.010$  for scalar).

After validating the measurement model, the structural model was tested using SEM. The hypothesized paths—representing associations among TTF, PU, and RGU, including an indirect association between TTF and RGU through PU—were estimated. Mediation was assessed with nonparametric bootstrapping (5,000 resamples) to generate bias-corrected 95% confidence intervals. Robustness checks included comparing a full mediation model with a partial mediation model and testing the influence of control variables such as gender, program level, and frequency of GenAI use.

## Results

### Descriptive analysis

Table 1 summarises the demographic profile of the 280 respondents. The sample was slightly female majority (43.4% male, 56.6% female) and concentrated in young adulthood, with over 90% of students aged between 18

Variable	Category	Frequency	Percent (%)
Gender	Male	122	43.4
	Female	159	56.6
Age	Under 18	3	1.1
	18–20	62	22.1
	21–23	57	20.3
	24–26	71	25.3
	27–29	63	22.4
	30 or above	25	8.9
Program level	Undergraduate	121	43.1
	Master	104	37.0
	Doctor	56	19.9
Frequency of using AI	Less than once a month	19	6.8
	1–3 Times a month	72	25.6
	1–2 Times a week	38	13.5
	3–4 Times a week	97	34.5
	Almost every day / Every day	55	19.6

**Table 1.** Demographic characteristics of the sample ( $n = 280$ ).

Item	N	Mean	SD	Min	Max
TTF1	280	3.268	1.128	1	5
TTF2	280	3.329	1.151	1	5
TTF3	280	3.357	1.186	1	5
TTF4	280	3.339	1.180	1	5
TTF5	280	3.275	1.167	1	5
PU1	280	3.339	1.171	1	5
PU2	280	3.300	1.121	1	5
PU3	280	3.318	1.131	1	5
PU4	280	3.325	1.151	1	5
RGU1	280	3.454	1.181	1	5
RGU2	280	3.429	1.121	1	5
RGU3	280	3.432	1.143	1	5
RGU4	280	3.296	1.095	1	5
RGU5	280	3.361	1.186	1	5
RGU6	280	3.354	1.182	1	5

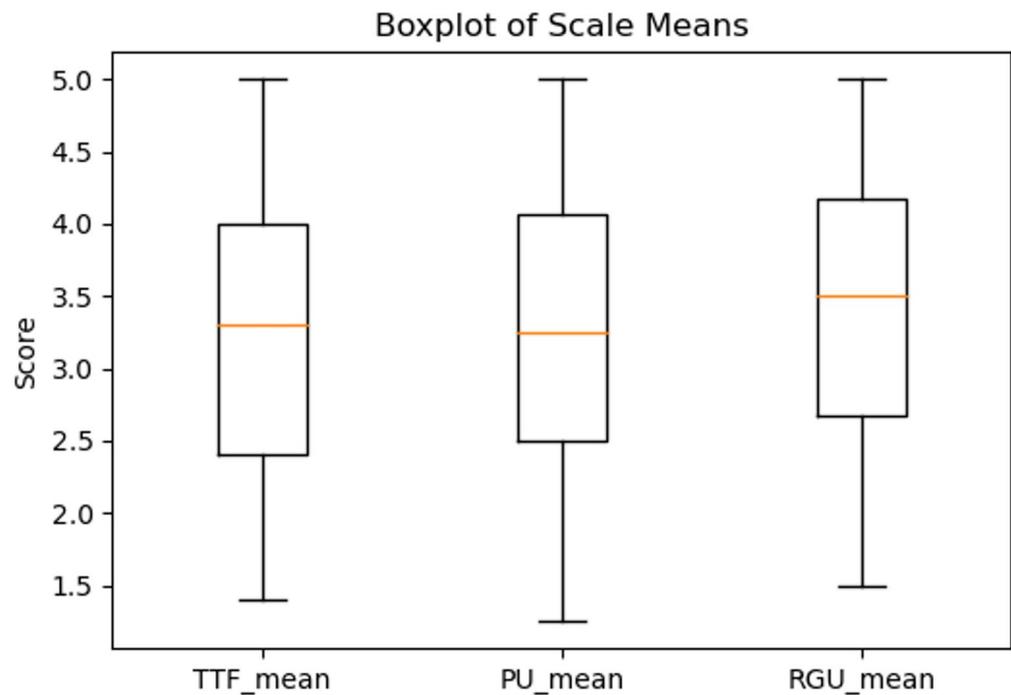
**Table 2.** Descriptive statistics for TTF, PU, and RGU items (Likert 1–5,  $n = 280$ ).

and 29. Around two fifths were undergraduates (43.1%), slightly more than one third were master's students (37.0%), and about one fifth were doctoral students (19.9%). With respect to GenAI use for academic purposes, most students reported regular use: more than half used GenAI at least once a week and nearly one fifth reported daily or almost daily use, whereas only 6.8% indicated very infrequent use (less than once a month). This profile indicates a diverse but GenAI-exposed student population that is suitable for examining perceptions of fit, usefulness, and responsibility.

Table 2 summarises the descriptive statistics for the items measuring task–technology fit (TTF), perceived usefulness (PU), and responsible GenAI use (RGU). Mean scores for all items ranged between 3.27 and 3.45, with standard deviations around 1.10–1.19, and responses spanned the full Likert range from 1 to 5. TTF items ranged from 3.27 to 3.36, PU items from 3.30 to 3.34, and RGU items from 3.29 to 3.45. As visualised in Fig. 2, the composite means cluster slightly above the scale midpoint with adequate interquartile spread, indicating neither floor nor ceiling effects and providing sufficient variability for reliability and structural analyses.

### Reliability and construct validity

Table 3 presents the results of the reliability analysis. Cronbach's  $\alpha$  values for TTF (0.879), PU (0.853), and RGU (0.909) all exceeded the recommended threshold of 0.70, indicating satisfactory internal consistency. In addition, the composite reliability (CR) values for the three constructs ranged from 0.901 to 0.929, further supporting the high reliability of the measurement model. The average variance extracted (AVE) values ranged from 0.674 to 0.694, all above the cut-off value of 0.50, thereby confirming adequate convergent validity.



**Fig. 2.** Distribution of TTF, PU, and RGU Scale Means (Boxplots; Likert 1–5,  $n=280$ ).

Construct	Cronbach's $\alpha$	CR	AVE
TTF	0.879	0.912	0.674
PU	0.853	0.901	0.694
RGU	0.909	0.929	0.687

**Table 3.** Reliability results for TTF, PU, and RGU.

	TTF	PU	RGU
TTF	0.821	0.314	0.373
PU	0.314	0.833	0.340
RGU	0.373	0.340	0.829

**Table 4.** Fornell–Larcker criterion.

Construct1	Construct2	HTMT
TTF	PU	0.363
TTF	RGU	0.415
PU	RGU	0.382

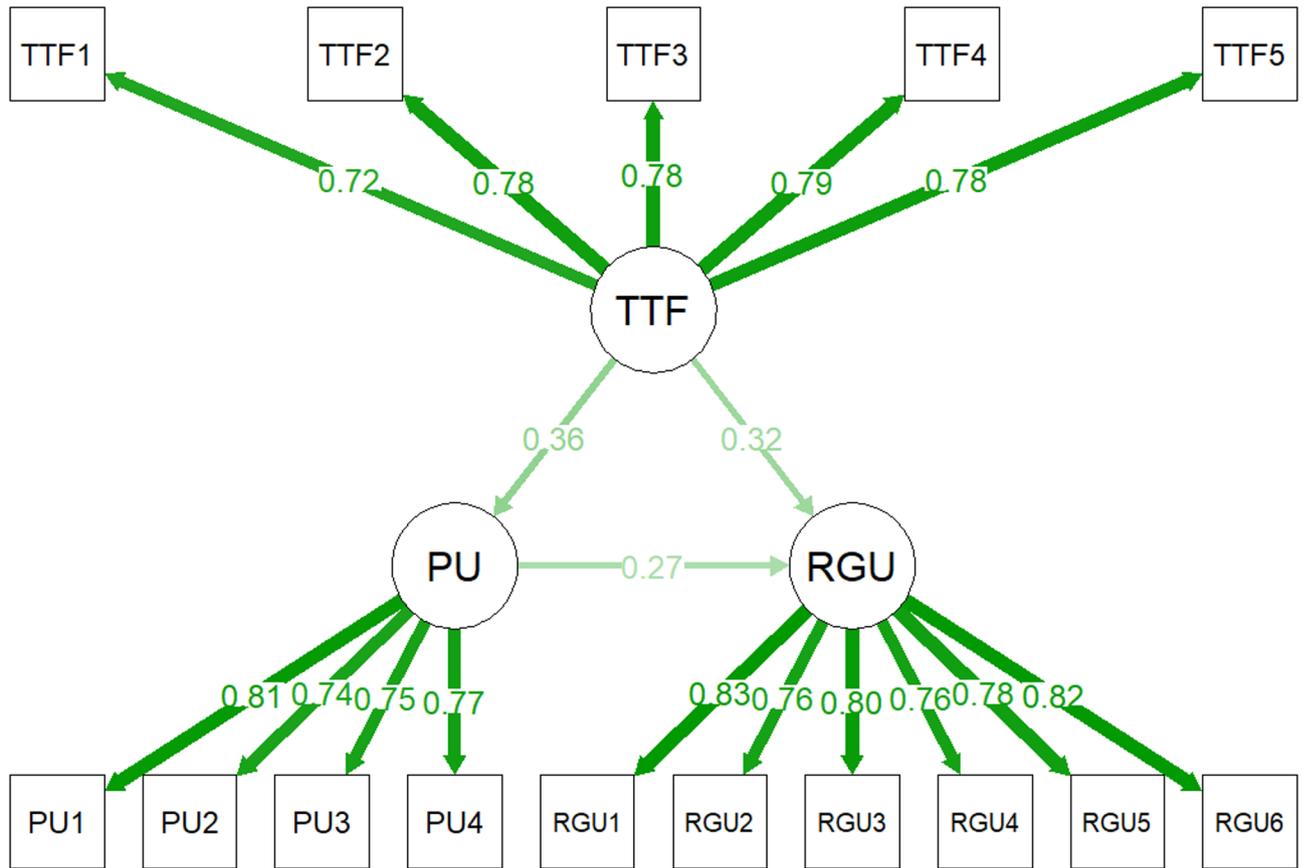
**Table 5.** HTMT results.

To assess discriminant validity, both the Fornell–Larcker criterion and the HTMT ratio were applied. As shown in Table 4, the square roots of AVE for TTF (0.821), PU (0.833), and RGU (0.829) were greater than the corresponding inter-construct correlations, satisfying the Fornell–Larcker criterion. Similarly, Table 5 reports the HTMT values, which ranged from 0.363 to 0.415. All values were well below the conservative threshold of 0.85, further confirming discriminant validity.

Taken together, these results demonstrate that the measurement model has strong reliability, convergent validity, and discriminant validity, providing a solid foundation for subsequent structural model analysis.

$\chi^2$	df	$\chi^2/df$	CFI	TLI	RMSEA	RMSEA 90% CI	SRMR
94.598	87	1.087	0.997	0.996	0.018	[0.000, 0.038]	0.027

**Table 6.** CFA model fit indices (ML).



**Fig. 3.** Structural Equation Model.

Dependent	Predictor	B	SE	z	p	$\beta$
PU	TTF	0.386	0.075	5.112	0.000	0.360
RGU	TTF	0.364	0.079	4.599	0.000	0.318
RGU	PU	0.292	0.074	3.943	0.000	0.273

**Table 7.** Structural paths (ML; unstandardized B and standardized  $\beta$ ).

**Structural model analysis**

The measurement and structural models were estimated using ML with FIML ( $n=280$ ). The three-factor measurement model exhibited excellent global fit (Table 6):  $\chi^2(87)=94.598$ ,  $\chi^2/df=1.087$ , CFI=0.997, TLI=0.996, RMSEA=0.018 with 90% CI [0.000, 0.038], and SRMR=0.027. The standardized factor structure and the hypothesized structural relations are visualized in Fig. 3, which also presents the standardized path estimates for interpretation alongside the fit indices in Table 6.

In the structural model, all hypothesized paths were positive and statistically significant (Table 7; see also Fig. 3). Task-Technology Fit (TTF) was positively associated with Perceived Usefulness (PU;  $\beta=0.360$ ,  $p<.001$ ), and both TTF ( $\beta=0.318$ ,  $p<.001$ ) and PU ( $\beta=0.273$ ,  $p<.001$ ) were positively related to Responsible GenAI Use (RGU). These results indicate that closer alignment between GenAI tools and academic tasks is associated with higher perceived usefulness and, in turn, with more responsible use behaviors, while a direct path from TTF to RGU remains (Table 7; Fig. 3).

Bootstrapped mediation analysis (5,000 draws; percentile CIs) further supported a partial-mediation pattern via PU (Table 8). The indirect effect of TTF on RGU through PU was  $B=0.113$ , 95% CI [0.055, 0.190],  $\beta=0.098$ ,

Effect	B	SE	z	p	95% CI (perc)	$\beta$
a: PU ~ TTF	0.386	0.076	5.061	0.000	[0.245, 0.544]	0.360
c': RGU ~ TTF	0.364	0.079	4.604	0.000	[0.218, 0.528]	0.318
b: RGU ~ PU	0.292	0.072	4.027	0.000	[0.157, 0.441]	0.273
indirect = a*b	0.113	0.034	3.279	0.001	[0.055, 0.190]	0.098
total = c' + a*b	0.476	0.075	6.353	0.000	[0.337, 0.633]	0.416

**Table 8.** Bootstrap mediation (ML; 5,000 draws; percentile 95% CI).

Group	Step	$\chi^2$	df	CFI	TLI	RMSEA	SRMR
Gender	Configural	201.25	174	0.988	0.985	0.033	0.039
Gender	Metric	212.586	186	0.988	0.986	0.032	0.047
Gender	Scalar	216.958	198	0.991	0.991	0.026	0.048
Program	Configural	283.556	261	0.99	0.988	0.03	0.046
Program	Metric	298.295	285	0.994	0.993	0.022	0.054
Program	Scalar	322.951	309	0.994	0.994	0.022	0.056
Frequency	Configural	589.651	435	0.919	0.903	0.080	0.078
Frequency	Metric	623.895	483	0.927	0.920	0.072	0.086
Frequency	Scalar	682.854	531	0.921	0.922	0.071	0.091

**Table 9.** Measurement invariance across groups: model fit indices.

Group	Step	$\Delta$ CFI	$\Delta$ RMSEA	$\Delta$ SRMR	Decision
Gender	Metric – configural	0	-0.001	0.008	Pass
Gender	Scalar – metric	0.003	-0.006	0.001	Pass
Program	Metric – configural	0.004	-0.008	0.008	Pass
Program	Scalar – metric	0	0	0.002	Pass
Frequency	Metric – configural	0.008	-0.008	0.008	Pass
Frequency	Scalar – metric	-0.006	-0.001	0.005	Pass

**Table 10.** Measurement invariance: change in fit indices ( $\Delta$ ) relative to the previous step.

$p = .001$ . The direct path from TTF to RGU remained significant ( $c'$ :  $\beta = 0.318$ ,  $p < .001$ ), and the total effect was  $B = 0.476$ , 95% CI [0.337, 0.633],  $\beta = 0.416$ ,  $p < .001$  (Table 8). Taken together, the evidence from the model fit (Table 6), structural paths (Table 7; Fig. 3), and mediation estimates (Table 8) indicates that TTF is positively related to responsible GenAI use both directly and indirectly via perceived usefulness.

### Group differences

As a prerequisite to subgroup comparisons, multi-group CFA established measurement invariance across gender, program level, and frequency of GenAI use. As shown in Table 9, the configural models demonstrated acceptable to excellent fit for all grouping variables. Changes in fit from configural to metric and from metric to scalar were within recommended thresholds for all three groupings (see Table 10 for  $\Delta$ CFI,  $\Delta$ RMSEA, and  $\Delta$ SRMR), supporting both metric and scalar invariance. These results justify the subsequent comparisons of composite means and the interpretation of observed group patterns.

Group comparisons were then conducted on the composite means of task–technology fit (TTF\_mean), perceived usefulness (PU\_mean), and responsible GenAI use (RGU\_mean). Descriptive patterns (Table 11) show that females reported higher averages than males on all three constructs; mean scores also increased steadily with more frequent GenAI use, and doctoral students tended to outscore undergraduates on TTF and PU.

For gender, Welch t-tests indicated no difference in TTF ( $t(256.91) = -0.573$ ,  $p = .567$ ), a marginal difference in PU favoring females ( $t(268.73) = -1.740$ ,  $p = .083$ ), and a significant difference in RGU with females reporting higher responsible use ( $t(273.29) = -3.192$ ,  $p = .0016$ ; Hedges'  $g = -0.377$ , 95% CI [-0.615, -0.130]) (Table 12).

Across program levels, Welch ANOVA showed significant between-group differences for TTF ( $F(2, 145.53) = 4.531$ ,  $p = .012$ ) and PU ( $F(2, 133.60) = 6.229$ ,  $p = .003$ ), but not for RGU ( $F(2, 138.23) = 0.042$ ,  $p = .959$ ) (Table 13). Games–Howell post hoc tests indicated that doctoral students scored higher than undergraduates on TTF (mean diff = 0.445,  $p = .009$ ;  $g = -0.487$ ), and that both master's and doctoral students scored higher than undergraduates on PU (Undergrad vs. Master: mean diff = 0.361,  $p = .006$ ,  $g = -0.411$ ; Undergrad vs. Doctoral: mean diff = 0.475,  $p = .023$ ,  $g = -0.465$ ) (Table 14).

Gender + Program level + Frequency				
Gender	N	TTF_mean	PU_mean	RGU_mean
Male	121	3.276	3.209	3.186
Female	159	3.342	3.406	3.541
Program level	N	TTF_mean	PU_mean	RGU_mean
Undergrad	121	3.180	3.093	3.369
Master	104	3.304	3.454	3.404
Doctoral	55	3.625	3.568	3.397
Frequency	N	TTF_mean	PU_mean	RGU_mean
Less than once a month	19	2.558	2.632	2.956
1–3 times a month	72	2.856	2.931	2.928
1–2 times a week	38	3.263	3.263	3.469
3–4 times a week	96	3.367	3.370	3.410
Almost every day / Every day	55	4.116	4.023	4.042

**Table 11.** Group-wise descriptive means of TTF, PU, and RGU ( $n=280$ ).

Outcome	group1	group2	n1	n2	t	df	p	95% CI (mean diff)	Hedges' g	95% CI (g)
TTF_mean	Male	Female	121	159	-0.573	256.905	0.567	[-0.293, 0.161]	-0.069	[-0.305, 0.167]
PU_mean	Male	Female	121	159	-1.740	268.733	0.083	[-0.420, 0.026]	-0.207	[-0.443, 0.030]
RGU_mean	Male	Female	121	159	-3.192	273.294	0.0016	[-0.574, -0.136]	-0.377	[-0.615, -0.13]

**Table 12.** Gender differences in TTF, PU, and RGU (Welch t-tests and hedges' g).

Outcome	df1	df2	F	p
TTF_mean	2	145.527	4.531	0.012
PU_mean	2	133.599	6.229	0.003
RGU_mean	2	138.234	0.042	0.959

**Table 13.** Program level differences: Welch ANOVA and.

Outcome	group1	group2	Mean diff	95% CI	p_adj	Hedges' g
TTF_mean	Undergrad	Doctoral	0.445	[0.094, 0.797]	0.009	-0.487
PU_mean	Undergrad	Master	0.361	[0.089, 0.633]	0.006	-0.411
PU_mean	Undergrad	Doctoral	0.475	[0.053, 0.897]	0.023	-0.465

**Table 14.** Program level differences: Games–Howell post hoc.

Outcome	df1	df2	F	p
TTF_mean	4	86.714	27.199	<0.001
PU_mean	4	84.766	15.095	<0.001
RGU_mean	4	86.880	15.565	<0.001

**Table 15.** Frequency-of-use differences: Welch ANOVA.

By frequency of GenAI use, Welch ANOVA detected robust differences for all three outcomes—TTF ( $F(4, 86.71) = 27.199, p < .001$ ), PU ( $F(4, 84.77) = 15.095, p < .001$ ), and RGU ( $F(4, 86.88) = 15.565, p < .001$ ) (Table 15). Games–Howell contrasts revealed a clear monotonic pattern in which more frequent users reported higher TTF, PU, and RGU; in particular, daily users outperformed infrequent users with large effects (e.g., vs. less than once per month: TTF mean diff = 1.559,  $p < .001$ ,  $g = -2.058$ ; PU mean diff = 1.391,  $p < .001$ ,  $g = -1.402$ ; RGU mean diff = 1.086,  $p < .001$ ,  $g = -1.323$ ), alongside several additional significant contrasts (Table 16). These very large standardized differences particularly reflect comparisons between the smallest, least frequent-user group ( $n = 19$ ) and daily users, where unequal cell sizes and variability can inflate Hedges' g. They should therefore

Outcome	group1	group2	Mean diff	p_adj	Hedges' g
TTF_mean	Less < 1/mo	1–2/wk	0.705	0.034	–0.807
TTF_mean	Less < 1/mo	3–4/wk	0.809	0.004	–0.874
TTF_mean	Less < 1/mo	Daily	1.559	< 0.001	–2.058
TTF_mean	1–3/mo	3–4/wk	0.511	0.001	–0.592
TTF_mean	1–3/mo	Daily	1.261	< 0.001	–1.704
TTF_mean	1–2/wk	Daily	0.853	< 0.001	–1.059
TTF_mean	3–4/wk	Daily	0.750	< 0.001	–0.858
PU_mean	Less < 1/mo	Daily	1.391	< 0.001	–1.402
PU_mean	1–3/mo	3–4/wk	0.439	0.005	–0.517
PU_mean	1–3/mo	Daily	1.092	< 0.001	–1.373
PU_mean	1–2/wk	Daily	0.760	< 0.001	–0.895
PU_mean	3–4/wk	Daily	0.653	< 0.001	–0.687
RGU_mean	Less < 1/mo	Daily	1.086	< 0.001	–1.323
RGU_mean	1–3/mo	1–2/wk	0.541	0.014	–0.634
RGU_mean	1–3/mo	3–4/wk	0.482	0.007	–0.522
RGU_mean	1–3/mo	Daily	1.114	< 0.001	–1.337
RGU_mean	1–2/wk	Daily	0.573	0.009	–0.722
RGU_mean	3–4/wk	Daily	0.633	< 0.001	–0.705

**Table 16.** Frequency-of-use differences: Games–Howell significant contrasts ( $p < .05$ ).

be interpreted as upper-bound indicators of strong directional differences, rather than as precise estimates of practical effect magnitude. Taken together, differences are most pronounced for usage frequency, modest for program level, and limited for gender, with the notable exception that females reported higher responsible use. In line with this, the subgroup patterns are best viewed as observational evidence of robust trends, particularly for usage frequency, rather than as finely calibrated estimates of the magnitude of practical differences.

## Discussion

This study examined how task–technology fit and perceived usefulness jointly relate to responsible generative-AI use in higher education and tested the model with covariance-based SEM. The measurement model showed excellent global fit, which supports the adequacy of the operationalization of the three latent constructs. Addressing RQ1 to RQ3, the structural results confirmed all hypothesized relationships. Task–Technology Fit (TTF) was positively associated with Perceived Usefulness (PU), and both TTF and PU were positively related to Responsible GenAI Use (RGU). Mediation estimates based on percentile bootstrap indicated that PU partially mediates the association between TTF and RGU, while the direct path from TTF to RGU remained significant. Beyond statistical significance, the reported  $f^2$  values indicate that these relationships are small to moderate in magnitude according to Cohen's (1988) guidelines, suggesting that TTF and PU make a meaningful, but not overwhelming, contribution to RGU. Together, these findings indicate that better alignment between GenAI functionalities and academic tasks is associated with higher usefulness perceptions and with more responsible use behaviors, and that alignment also carries an independent contribution to responsibility beyond usefulness.

Placed in the context of prior models, the positive TTF to PU link is consistent with the central logic of Task–Technology Fit theory, which argues that performance benefits arise when technology characteristics match task requirements<sup>16</sup>. The relationship between PU and downstream outcomes aligns with the Technology Acceptance Model, which positions perceived usefulness as a key determinant of acceptance and use<sup>12</sup>. The present study contributes to this literature by positioning responsibility as an outcome adjacent to adoption and performance. Using items adapted from recent work on responsible GenAI behaviors in academic settings<sup>1</sup>, the results show that when GenAI is both well fitted to tasks and viewed as useful, students report more disclosure, policy alignment, and integrity-consistent practices. The persistence of a direct TTF to RGU path after controlling for PU suggests that fit may also provide normative clarity about appropriate boundaries of AI assistance, which can be associated with more responsible conduct even when perceived benefits are held constant.

Turning to RQ4, the exploratory subgroup analyses provide contextual evidence on how responsibility patterns vary across student groups. Descriptive means indicate higher scores for females and for students who use GenAI more frequently, and higher TTF and PU among doctoral students compared with undergraduates. Inferential tests show a significant gender difference only for RGU favoring females, significant program-level differences for TTF and PU but not RGU, and robust frequency-of-use differences for all three constructs with a clear monotonic pattern in which daily users outperform less frequent users. One plausible interpretation is that female students may engage more strongly with institutional guidelines, disclosure norms, and risk-avoidant strategies in academic settings, which is reflected in higher responsible-use scores even when perceptions of fit and usefulness are broadly similar. Likewise, the higher TTF and PU among doctoral students likely reflect their involvement in more complex and research-intensive tasks, for which GenAI tools provide clearer performance benefits and closer alignment with academic demands than for routine undergraduate coursework. These patterns suggest that opportunities to practice with GenAI may be related to differences in both perceived fit

and responsible behaviors, complementing the structural relations reported above. The higher RGU reported by female students is consistent with prior work that links women to greater concern for academic integrity and higher risk sensitivity in educational settings, which may encourage more cautious, transparent engagement with GenAI. The higher TTF and PU among doctoral students likely reflect the fact that their academic work involves more complex research, writing, and data analysis tasks, for which GenAI tools provide clearer performance benefits and closer task alignment than for routine undergraduate coursework. Measurement invariance analyses indicated that the measurement structure was comparable across gender and program level, and although baseline fit was weaker for usage-frequency groups, the change-in-fit criteria were met; this strengthens the interpretability of the subgroup patterns reported in Tables 9, 10, 11, 12, 13, 14, 15 and 16 while warranting modest caution for frequency-based contrasts.

The findings have clear practical implications for instructors, institutions, and training programmes. Because both TTF and PU showed statistically significant, small-to-moderate effects on RGU, instructional interventions that enhance perceived fit and usefulness are likely to produce meaningful improvements in responsible GenAI use. For instructors designing GenAI-integrated coursework, activities should be explicitly tied to identifiable steps of academic tasks. Rubrics can specify allowed and disallowed uses, and assignment prompts can ask students to explain how they used GenAI at each step. To strengthen the TTF-to-PU link while supporting responsibility, usefulness should be made visible within clear boundaries by providing concrete examples of value, brief guidance on citing and verifying AI output, and simple disclosure statements that students can insert into submissions. These measures reinforce the PU-to-RGU pathway and help translate favourable evaluations into responsible practice.

At the institutional level, universities can embed brief AI-use policies and disclosure templates in course syllabi and learning management systems, providing default wording for permissible and impermissible uses and simple reporting channels for uncertainty about responsible use. Short, scenario-based workshops on responsible AI literacy can complement these structural supports by walking students through typical GenAI-integrated tasks, examples of appropriate and inappropriate assistance, and hands-on exercises in verifying and documenting AI output. The subgroup analyses showed consistently higher means for all three constructs among daily users and higher RGU among female students, indicating that institutional policies should prioritise support for less frequent users and undergraduates while recognising the potential of high-RGU groups as role models. Guided by the RQ4 results, support can be differentiated for less frequent users and for undergraduates. Peer exemplars, particularly female students who reported higher levels of RGU, can also be leveraged to diffuse responsible practices across student communities.

Several considerations qualify the conclusions and point to future work. First, the study relied exclusively on self-report questionnaire measures and a cross-sectional survey design. This combination limits causal inference and makes the results susceptible to common-method variance and social desirability bias, particularly for items that explicitly reference “responsible” behaviours. Because RGU involves disclosure of ethically desirable practices (e.g., policy compliance, verification), students may have overstated their level of responsibility, even though the anonymity of the survey was emphasized. As in other survey-based research on ethics and academic integrity, self-reported measures of responsible behaviour may not fully align with students’ actual practices in coursework, so the RGU scores should be interpreted as reflecting perceived tendencies and intentions rather than directly observed behaviour. Future research could combine self-reports with behavioural indicators, such as actual disclosure statements in assignments or logged interactions with GenAI tools, to cross-validate self-reported responsibility against observable practices and thereby mitigate this source of bias.

Second, the data were collected from a China-only sample of higher-education students using purposive, non-probability sampling and including only students with prior experience using GenAI for academic purposes. This focus on current users was appropriate for measuring perceptions of task–technology fit, usefulness, and responsible practices, but it also introduces self selection bias and excludes a potentially important cohort of non users. Students who avoid GenAI because of ethical concerns, fear of misuse, limited digital literacy, or low technology readiness may hold systematically different attitudes toward GenAI, may perceive lower task–technology fit and usefulness, and may conceptualise responsible use in different ways. In addition, the cultural and institutional characteristics of this Chinese university context, including local norms surrounding academic integrity and rapidly evolving GenAI policies, may differ from those in other national systems and thus limit the cross-cultural generalisability of the findings. As a result, the levels of TTF, PU, and RGU observed in this study are likely to characterise current or emerging GenAI users rather than the broader student population, and future research should include both users and non users in order to examine how responsibility patterns differ across these groups and to improve the generalisability of the findings. In addition, the study did not collect detailed information on participants’ specific degree programmes or courses, which limits the ability to examine discipline-level differences and further constrains the generalisability of the findings across different curricular contexts. Future work should therefore record and compare students’ programme and course characteristics to identify whether task–technology fit, perceived usefulness, and responsible GenAI use vary systematically across subject domains.

Third, although measurement-invariance analyses indicated that the measurement structure was comparable across gender and program level, and change-in-fit criteria were met for usage-frequency groups, the relatively weaker baseline fit for the frequency models (e.g., RMSEA = 0.080 for the configural model) suggests that frequency-based contrasts should be interpreted with modest caution. These subgroup patterns are therefore best viewed as tentative and in need of confirmation in larger and more heterogeneous samples.

Future studies could use longitudinal or experimental designs that manipulate task alignment to test causal effects on usefulness and responsibility, incorporate behavioral indicators of disclosure and verification in coursework, and evaluate the stability of the measurement and structural relations through multi-group analyses. Despite these limitations, the convergent evidence from model fit, structural paths, and mediation

estimates provides a coherent account that integrates TTF and TAM perspectives with responsible GenAI use. The results suggest that aligning tools with tasks and making benefits explicit within clear boundaries are pivotal not only for effectiveness but also for sustainable and integrity-consistent use in higher education. The subgroup contrasts reported for RQ4 should be interpreted as observational patterns rather than causal effects, particularly given unequal cell sizes and potential unobserved confounds. Consistent with the SAGER recommendations, the gender-based comparisons are understood as reflecting sociocultural and educational differences in GenAI exposure and norms rather than fixed biological characteristics.

## Conclusion

The study provides empirical evidence that task–technology fit and perceived usefulness are jointly associated with responsible GenAI use among higher-education students. The measurement model exhibited excellent fit, and the structural model indicated that TTF was positively associated with PU, and that both TTF and PU were positively related to RGU. Mediation analyses further indicated a partial mediation pattern in which PU was associated with the link between TTF and RGU, suggesting that responsibility may be related both to clearer task alignment and to stronger usefulness perceptions. Supplementary exploratory subgroup analyses suggested systematic differences across gender, program level, and frequency of use, with the steepest gradients for more frequent users; these observational patterns complement the structural results but warrant cautious interpretation.

These findings advance adoption theory by linking ethical engagement with core adoption constructs and offer actionable guidance for practice: design task-aligned GenAI activities, scaffold usefulness within boundaries, and institutionalize transparent disclosure and micro-training. In implementation, practitioners can prioritize targeted supports for less frequent users and early-stage learners, provide concise disclosure templates and verification checklists, and leverage peer exemplars to normalize responsible practices. Future work should employ longitudinal or experimental designs, multi-group tests, and behavioural indicators of disclosure and verification to strengthen robustness and allow more rigorous causal inference. As higher education navigates GenAI integration, aligning tools with tasks and highlighting authentic academic value appear pivotal not only for effectiveness but also for sustainable, integrity-consistent use.

## Data availability

The data presented in this study are openly available in Zenodo at <https://doi.org/10.5281/zenodo.17157694>, reference number 17157694.

Received: 23 October 2025; Accepted: 11 December 2025

Published online: 02 January 2026

## References

- Al Maharmah, A., Al-Qaysi, N. & Al-Emran, M. Development and validation of the AI academic integrity scale (AAIS). *Comput. Educ.* **210**, 105084. <https://doi.org/10.1016/j.compedu.2025.105084> (2025).
- Al Zaidy, A. The impact of generative AI on student engagement and ethics in higher education. *J. Inform. Technol. Cybersecur. Artif. Intell.* **1** (1), 30–38. <https://doi.org/10.70715/jitcai.2024.v1.i1.004> (2024).
- Alamir, I. A. Extended technology acceptance model (TAM) for evaluating digital banking adoption in Saudi Arabia. *Sci. J. King Faisal University: Humanit. Manage. Sci.* **26** (1), 63–70. <https://doi.org/10.37575/h/mng/240043> (2025).
- Al-Dokhny, A. et al. Can multimodal large Language models enhance performance benefits among higher education students? An investigation based on the task–technology fit theory and the artificial intelligence device use acceptance model. *Sustainability* **16** (23), 10780. <https://doi.org/10.3390/su162310780> (2024).
- Aldreabi, M., Khan, M. & Alenezi, M. Determinants of student adoption of generative AI in higher education: A modified UTAUT2 approach. *Electron. J. e-Learning.* **23** (2), 110–127 (2025). <https://academic-publishing.org/index.php/ejel/article/view/3599>
- Al-Mamary, Y. H., Alfalah, A. A., Alshammari, M. M. & Abubakar, A. A. Exploring factors influencing university students' intentions to use chatgpt: analysing task-technology fit theory to enhance behavioural intentions in higher education. *Future Bus. J.* **10** (1), 119. <https://doi.org/10.1186/s43093-024-00406-5> (2024).
- Alyoussef, I. Y. E-learning acceptance: the role of task–technology fit as sustainability in higher education. *Sustainability* **13** (11), 6450. <https://doi.org/10.3390/su13116450> (2021).
- Atkinson-Toal, A. & Guo, C. Generative artificial intelligence (AI) education policies of UK universities. *Enhancing Teach. Learn. High. Educ.* **2**, 70–94. <https://doi.org/10.62512/ethhe.20> (2024).
- Aurangzeb, W., Kashan, S. & Rehman, Z. U. Investigating technology perceptions among secondary school teachers: A systematic literature review on perceived usefulness and ease of use. *Acad. Educ. Social Sci. Rev.* **4** (2), 160–173. <https://doi.org/10.48112/aessr.v4i2.746> (2024).
- Chan, C. K. Y. & Lee, K. K. W. The AI generation gap: are gen Z students more interested in adopting generative AI such as ChatGPT in teaching and learning than their gen X and millennial generation teachers? *Smart Learn. Environ.* **10**, 23. <https://doi.org/10.1186/s40561-023-00269-3> (2023).
- Dabis, A. & Csáki, C. AI and ethics: investigating the first policy responses of higher education institutions to the challenge of generative AI. *Humanit. Social Sci. Commun.* **11**, 1006. <https://doi.org/10.1057/s41599-024-03526-z> (2024).
- Davis, F. D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q.* **13** (3), 319–340. <https://doi.org/10.2307/249008> (1989).
- Foung, D., Lin, L. & Chen, J. Reinventing assessments with ChatGPT and other online tools: opportunities for GenAI-empowered assessment practices. *Computers Education: Artif. Intell.* **6**, 100250. <https://doi.org/10.1016/j.caeai.2024.100250> (2024).
- Geddam, S. M., Nethravathi, N. & Hussian, A. A. Understanding AI adoption: the mediating role of attitude in user acceptance. *J. Inf. Educ. Res.* **4** (2), 1664–1672 (2024). <http://jier.org>
- Gesser-Edelsburg, A., Hijazi, R., Eliyahu, E. & Tal, A. Bridging the divide: an empirical investigation of artificial intelligence and generative artificial intelligence integration across genders, disciplines and academic roles. *Eur. J. Open. Distance E-Learning.* **26** (s1), 51–69. <https://doi.org/10.2478/eurodl-2024-0008> (2024).
- Goodhue, D. L. & Thompson, R. L. Task-technology fit and individual performance. *MIS Q.* **19** (2), 213–236. <https://doi.org/10.2307/249689> (1995).

17. Gutiw, D., Sorg, J. M. & Cruz Rodriguez, G. Responsible use of artificial intelligence: Perspective of a global IT management consultancy. In S. R. Islam, A. R. Al-Ali, & T. A. Gulliver (Eds.), *Cases on AI ethics in business* (pp. 15–30). IGI Global. (2024). <https://doi.org/10.4018/979-8-3693-2643-5.ch010>
18. Hair, J. F. Jr., Black, W. C., Babin, B. J. & Anderson, R. E. *Multivariate data analysis* (8th ed.). Cengage. (2019).
19. Henseler, J., Ringle, C. M. & Sarstedt, M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. *J. Acad. Mark. Sci.* **43** (1), 115–135. <https://doi.org/10.1007/s11747-014-0403-8> (2015).
20. Hou, I., Nguyen, H. V., Man, O. & MacNeil, S. The evolving usage of GenAI by computing students. In *Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 2 (SIGCSE TS 2025)* (p. 2). ACM. (2025). <https://doi.org/10.1145/364155.5.3705266>
21. Hu, L. T. & Bentler, P. M. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Model.* **6** (1), 1–55. <https://doi.org/10.1080/10705519909540118> (1999).
22. Jebri, H. A. M., Mohamad, U. H. & Ahmad, M. N. Exploring the application of ISSM and TTF theories: A literature review. *Int. J. Adv. Sci. Eng. Inform. Technol.* **14** (5), 1589–1595. <https://doi.org/10.18517/ijaseit.14.5.19171> (2024).
23. Kline, R. B. *Principles and Practice of Structural Equation Modeling* 4th edn (Guilford Press, 2016).
24. Kotsis, K. T. Artificial intelligence creates plagiarism or academic research? *Eur. J. Arts Humanit. Social Sci.* **1** (6), 169–179. [https://doi.org/10.59324/ejahss.2024.1\(6\).18](https://doi.org/10.59324/ejahss.2024.1(6).18) (2024).
25. Liu, Y., Park, J. & McMinn, S. Using generative artificial intelligence/ChatGPT for academic communication: students' perspectives. *Int. J. Appl. Linguistics.* **34** (4), 1437–1461. <https://doi.org/10.1111/ijal.12574> (2024).
26. Marín, Y. R. et al. Ethical challenges associated with the use of artificial intelligence in university education. *J. Acad. Ethics.* <https://doi.org/10.1007/s10805-025-09660-w> (2025). Advance online publication.
27. Meakin, L. A. Embracing generative AI in the classroom whilst being mindful of academic integrity. In S. Mahmud (Ed.), *Academic integrity in the age of artificial intelligence* (58–77). IGI Global. (2024). <https://doi.org/10.4018/979-8-3693-0240-8.ch004>
28. Nunnally, J. C. & Bernstein, I. H. *Psychometric Theory* 3rd edn (McGraw-Hill, 1994).
29. Oktaria, S. D., Firdaus, R. & Pangestu, D. Faktor-faktor yang mempengaruhi penerimaan teknologi dalam pembelajaran di SMA Bandar Lampung menggunakan model TAM di era disrupsi. In *Proceedings of the National Seminar on Technology, Local Wisdom, and Transformative Education (SNTEKAD)*, Universitas Ahmad Dahlan. 2(1) 139–148. <https://doi.org/10.12928/sntekad.v2i1.19034> (2025).
30. Sergeeva, O. V. et al. Understanding higher education students' adoption of generative AI technologies: an empirical investigation using UTAUT2. *Contemp. Educational Technol.* **17** (2), ep571. <https://doi.org/10.30935/cedtech/16039> (2025).
31. Sodik, S. & Rokib, M. Indonesian students' use of chat generative Pre-trained transformer in essay writing practices. *Int. J. Evaluation Res. Educ.* **13** (4), 2698–2706. <https://doi.org/10.11591/ijere.v13i4.28956> (2024).
32. Sousa, A. E. & Cardoso, P. Use of generative AI by higher education students. *Electronics* **14** (7), 1258. <https://doi.org/10.3390/electronics14071258> (2025).
33. Toros, E., Asiksoy, G. & Sürücü, L. Refreshment students' perceived usefulness and attitudes towards using technology: A moderated mediation model. *Humanit. Social Sci. Commun.* **11**, 333. <https://doi.org/10.1057/s41599-024-02839-3> (2024).
34. Ullah, M., Naeem, B., Kamel Boulos, M. N. & S., & Assessing the guidelines on the use of generative artificial intelligence tools in universities: A survey of the world's top 50 universities. *Big Data Cogn. Comput.* **8** (12), 194. <https://doi.org/10.3390/bdcc8120194> (2024).
35. Wahle, J. P., Ruas, T., Mohammad, S. M., Meuschke, N. & Gipp, B. AI usage cards: Responsibly reporting AI-generated content. *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries (JCDL 2023)* (1–12). ACM/IEEE. (2023). <https://doi.org/10.1145/3531146.3533231>

## Acknowledgements

The authors appreciate the administrative and technical support provided by the participating institutions. We also acknowledge the assistance of an online survey platform for questionnaire deployment. During the preparation of this manuscript, the authors used an AI-based language model to assist with language polishing and wording refinement, and a grammar-checking tool to ensure linguistic accuracy. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

## Author contributions

Q.W. conceived and designed the study, developed the research framework, collected and analyzed the data, and drafted the main manuscript text. M.A. provided theoretical guidance, contributed to model validation and interpretation, and critically revised the manuscript. Y.W. assisted with data acquisition, ethical approval, and language editing. All authors discussed the results, contributed to the final interpretation, and reviewed and approved the manuscript.

## Funding

The Authors received NO FUNDING for this work.

## Declarations

## Competing interests

The authors declare no competing interests.

## Ethics approval and informed consent

This study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Review Committee of Wuhan University of Engineering Science (Approval No. WUOES20250701, dated 1 July 2025). Written informed consent was obtained from all participants prior to data collection. Participation was voluntary, and all responses were anonymized to ensure confidentiality.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-32692-6>.

**Correspondence** and requests for materials should be addressed to Y.W.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026