Computers and Education Open 10 (2026) 100323

Contents lists available at ScienceDirect

Computers and Education Open

journal homepage: www.sciencedirect.com/journal/computers-and-education-open

Ethical conditions for university students' adoption of large language models in exam preparation contexts

Antonio Pérez-Portabella [a,c], Mario Arias-Oliva [b], Graciela Padilla-Castillo [c], Jorge de Andrés-Sánchez [d,*]

[a] Universitat Rovira i Virgili, Estudis de Comunicació, y Universidad Complutense de Madrid, Spain
[b] Marketing Department, University Complutense of Madrid, Faculty of Business & Economy; Campus de Somosaguas, 28223 Madrid, Spain
[c] Departamento de Periodismo III, Facultad de Ciencias de la Información, Universidad Complutense de Madrid, Spain
[d] Social and Business Research Lab., University Rovira i Virgili, Campus de Bellissens, 43204 Reus, Spain

ARTICLE INFO

Keywords:
Large language models (LLMs)
Multiple ethics scale (MES)
Higher education
Necessary condition analysis (NCA)
Partial Least Squares Structural Equation Modeling (PLS-SEM)

ABSTRACT

Large Language Models (LLMs), have become a disruptive force in higher education, assisting students in tasks such as summarizing content, clarifying concepts, and preparing for exams. Their rapid adoption has generated both enthusiasm and ethical concern, particularly regarding fairness and responsible use in assessment contexts. This study examines whether university students' ethical perceptions regarding the use of LLMs for exam preparation influence their intention to use them (IU), and whether this intention, in turn, affects their actual use (USE). Data were collected from 151 undergraduate students enrolled in Spanish universities. Ethical perceptions were measured using the Multidimensional Ethics Scale (MES), which captures three normative perspectives: moral equity (ME), consequentialism (CO), and deontology (DE). The first research objective (RO1) was to determine whether a minimum threshold in each ethical dimension is required for students to form the intention to use LLMs. The second (RO2) assessed whether higher ethical evaluations lead to greater acceptance and use. Necessary Condition Analysis revealed that all three ethical dimensions are necessary conditions for IU (ME: $d = 0.338$; CO: $d = 0.274$; DE: $d = 0.207$; all $p < 0.001$). Partial Least Squares-Structural Equation Modeling showed that only CO ($\beta = 0.350$, $p < 0.001$) and DE ($\beta = 0.329$, $p < 0.001$) exert statistically significant and sufficient effects on IU, while ME does not ($\beta = 0.108$, $p = 0.414$). Finally, IU is both a necessary ($d = 0.325$, $p < 0.001$) and sufficient predictor of USE ($\beta = 0.905$, $p < 0.001$). The findings highlight that ethical reasoning is central to the responsible adoption of LLMs in exam preparation, offering practical guidance for educators and institutions promoting academic integrity.

1. Introduction

The impact of artificial intelligence (AI) is especially significant in the field of education, and particularly in the university [1]. Large language models (LLMs), such as ChatGPT, have opened up new possibilities in teaching and learning processes [2], enabling everything from the automation of routine tasks to personalized learning. These tools offer immediate responses to complex questions, support academic writing, generate summaries, and even provide specialized virtual tutoring [3]. This technological revolution is not only redefining the role of the student as a knowledge recipient but also reshaping the role of educators, who can now rely on intelligent systems to design materials, identify learning needs, or conduct more accurate assessments [2]. The are several reports demonstrating that the collaboration between educators and LLMs, through reciprocal feedback review, substantially improves both the technical depth and practical relevance of the support provided to students [4,5].

However, the use of LLMs also raises critical questions, particularly regarding ethics, equity, and academic integrity [6]. Unsupervised or improper utilization can yield adverse outcomes, potentially undermining the acquisition of essential competencies such as written communication [7] and impeding the enhancement of advanced cognitive processes, including critical thinking, logical reasoning, and creative problem-solving [8].

* Corresponding author.
E-mail addresses: antonio.perezportabella@urv.cat (A. Pérez-Portabella), mario.arias@ucm.es (M. Arias-Oliva), gracielp@ucm.es (G. Padilla-Castillo), jorge.deandres@urv.cat (J. de Andrés-Sánchez).

https://doi.org/10.1016/j.caeo.2025.100323
Received 1 August 2025; Received in revised form 8 October 2025; Accepted 12 December 2025
Available online 13 December 2025

The presence of algorithmic biases in LLMs and other AI systems is a major concern. The outputs generated by these tools may reproduce and even amplify stereotypes or provide unbalanced information [9]. Furthermore, the inappropriate use of AI tools in academic assessment contexts—such as automated essay writing or unsupervised exam solving—raises serious questions about fairness in evaluating student performance [10], which has led many universities to explicitly classify their use in these contexts as misconduct [11]. Educators now face growing challenges in distinguishing between original student work and AI-generated content [3], which necessitates a reevaluation of traditional assessment methods and a move toward authentic learning approaches that transcend rote repetition or copying [12]. In addition, there is a disconnect between political regulation and practical implementation, as the responsible integration of AI in higher education currently lacks a clear national strategy and relies heavily on individual instructor's initiatives [13].

These considerations provide the motivation for the present research, which focuses on how university students' ethical judgments influence their effective use of LLMs as a resource for preparing for supervised exams. These exams constitute the main form of assessment in Spanish higher education [14]. In that study, based on the analysis of >1000 course syllabi, it was observed that nearly 70 % of subjects include a final written exam, which accounts for approximately 60 % of the total grade. The most common formats are multiple-choice and short-answer questions, while problem-solving and essay-based assessments are less frequent. Although some courses incorporate midterm exams (38.8 %) or complementary practical activities, the final exam still reflects a predominantly summative and traditional approach, focusing more on measuring outcomes than on learning processes. In this regard, San Martín et al. [15] noted that social science programs tend to rely more heavily on final exams (around 90 % of the syllabi reviewed), whereas technical degrees more often include partial examinations throughout the semester.

This study does not debate whether the use of LLMs to prepare exams violates academic norms. From a formalist perspective, their use is legally permitted in many academic activities, such as independent study. In fact, it may often be desirable, as students who employ self-regulated learning strategies tend to perform better academically [16], and LLMs can serve as excellent tools for individualized instruction [2]. In the Spanish context, although academic programs prescribe certain official materials for course follow-up, and instructors may recommend complementary resources [17], students retain the freedom to use the resources they deem most appropriate. These may include traditional tutoring [18] as well as online resources such as video tutorials, e-books, or webinars [19].

This study approaches the suitability of LLMs from an interpretive perspective, as students often construct their own understanding of what constitutes acceptable behavior, based on observational learning and social cues from instructors [20]. Interpretive rules are shaped within the informal dynamics of the classroom, influenced by social norms, values, and interactions that—although not always explicit—significantly affect student conduct and learning [21].

The approach used in this study is based on an adaptation of the Multidimensional Ethics Scale (MES) [22]. This scale has proven to be a particularly effective instrument for explaining decision-making with from an ethical perspective a wide range of domains [23] including educational contexts [24–29].

Fig. 1 depicts the theoretical framework of the study, which integrates the three ethical dimensions of MES—moral equity (ME), consequentialism (CO), and deontology (DE)—as antecedents of intention to use (IU) LLMs. In this context, IU is understood as the predisposition to use the technology under study [30]. In line with established technology acceptance models such as the Technology Acceptance Model (TAM) [30] and the Unified Theory of Acceptance and Use of Technology (UTAUT) [31], IU is posited as the direct antecedent of actual use (USE).

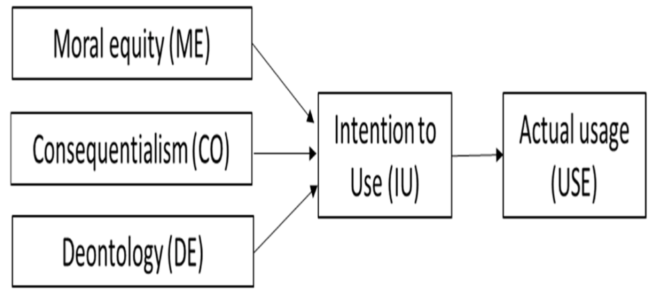

Fig. 1. Theoretical groundwork used in this paper. Source: Authors' own work by using Shawver and Sennetti [32].

Based on this framework, two research objectives (RO) are proposed, aimed at understanding how these ethical perceptions influence students' decisions to use LLMs as a study resource for preparing for supervised exams:

RO1: To determine whether the ethical dimensions included in the MES constitute necessary conditions for a student to develop an intention to use LLMs as an academic preparation tool—and, subsequently, whether such intention is necessary for actual use. Logically, this entails the following questions: Is a minimum level of a particular ethical perception (e.g., moral equity) required to reach a certain level of intention to use? Is a minimum level of intention required for actual use to occur?

RO2: To examine the extent to which these ethical dimensions act as sufficient conditions to explain LLM adoption in exam preparation. From a causal logic standpoint, this involves asking whether "a more positive ethical perception of dimension X leads to a higher level of acceptance" [33].

Although these two objectives require distinct analytical approaches [34], they are conceptually complementary [35]. The first is addressed through Necessary Condition Analysis (NCA), based on the methodological foundations developed by Dul [34] and Dul et al. [36], while the second is examined using a Partial Least Squares Structural Equation Modeling (PLS-SEM) approach, which is appropriate for this second research objective [35].

The operationalization of the significance testing of relationships between variables is relatively similar in both approaches. In PLS-SEM, as in any other regression-based method, the direction and strength of the relationships are represented by a coefficient, and their statistical significance is assessed through a p-value. In NCA, by contrast, the degree to which a variable is necessary is quantified through the necessity effect size (d), to which a p-value is assigned by means of a permutation test [36]. Within the NCA framework, an effect size of $d > 0.1$ is typically considered relevant, and—as in conventional statistical testing—a p-value below the 5 % significance level is required to support the hypothesis under examination.

2. Theoretical groundwork

2.1. Evaluation of acceptance and use

Following classical technology acceptance models such as TAM and UTAUT, intention to use (IU) is employed as a latent variable to measure the acceptance of LLMs. IU is defined as an individual's predisposition to perform a specific behavior [37], and in this study, it refers to university students' willingness to use LLMs as a tool for exam preparation—one of the most common uses reported by students in both qualitative [38,39] and quantitative studies [1], as well as in systematic reviews [40].

A straightforward use of LLMs is as a personal tutor, helping students to understand new concepts. As illustrated in the qualitative study by Leite [39], one participant stated, "*AI helped me to have a better understanding of theoretical concepts and to progress quickly in areas that were* *difficult for me*." However, the use of LLMs extends beyond tutoring: they can summarize and outline documents and even translate them when these are not in the student's native language [41] and logical–mathematical reasoning, such as writing code or solving mathematical problems [42]. This helps explain why previous studies have reported that students' primary use of LLMs is as an assistant for exam preparation [1].

To explain IU, a modified version of the MES developed by Shawver and Sennetti [32] was used. Modeling decision-making from an ethical standpoint requires acknowledging that distinguishing between ethical and unethical behavior in a unidimensional manner is overly simplistic, as moral philosophy offers multiple normative theories concerning the morality of actions [43].

As Fig. 1 shows, the MES incorporates latent variables derived from different normative theories, allowing for the coexistence of multiple ethical perspectives in the decision-making process, with the relevance of each depending on the specific context [22]. These perspectives range from those aligned with virtue ethics, which emphasize moral qualities such as goodness and justice—captured by the notion of moral equity—to those focused on the outcomes of actions, as in consequentialism, and on the moral duties that guide behavior, as in deontology. The MES model used in this study groups these ethical constructs into the three factors of moral equity, consequentialism, and deontology [32]. This theoretical pluralism justifies the application of the MES for analyzing ethical decision-making in academic and educational contexts (see, e.g., [26,28]).

Although IU is a strong predictor of actual use (USE), they are not equivalent. Various contextual factors may lead to discrepancies between intention and behavior [44]. In TAM, TAM2, and UTAUT models [30,31,45], it is assumed that the effective use of a technology is mediated by the perception of its advantages over existing alternatives.

The use of an LLM as a study tool competes with various alternatives, both technology-based [19] and traditional (e.g., textbooks, handwritten notes). Therefore, for effective use to occur, students must first evaluate the advantages of the new technology compared to existing ones [44], leading to a judgment of its acceptability. If this level of acceptance does not exceed a certain threshold, students are likely to opt for conventional methods they used prior to the emergence of LLMs. Conversely, once this threshold is surpassed, greater intention to use should translate into more frequent or intensive use. Accordingly, we propose:

Hypothesis 1a. (H1a): Intention to use LLMs for exam preparation is a necessary condition for actual use.

Hypothesis 1b. (H1b): Intention to use LLMs for exam preparation is positively related to actual use.

2.2. Moral equity

Moral equity (ME) refers to judgments about fairness, justice, and fundamental ethical principles such as equality and individual rights [26,46]. From the capabilities approach, justice is assessed not only in terms of resources or formal freedoms but also in terms of individuals' real opportunities to flourish [47]. ME, which reflects perceptions of fairness, honesty, and justice, is conceptually aligned with virtue ethics. From a virtue ethics perspective, moral action arises from the character of the individual and the pursuit of moral excellence [48]. Thus, ME can be viewed as an applied manifestation of virtue ethics in educational decision-making, where students evaluate the fairness and integrity of using LLMs as part of their learning process [22].

Within this framework, LLMs may enhance educational equity by serving as compensatory tools that help mitigate structural disadvantages. Their accessibility enables them to function as personal tutors for students who lack access to private lessons or specialized resources [49], allowing for personalized learning [12] tailored to diverse contexts—thus aligning with the capabilities perspective. Moreover, free or low-cost options such as ChatGPT 3.5 or DeepSeek promote adoption and help reduce economic barriers, supporting distributive justice [50].

LLMs also pose potential challenges from the perspective of ME, particularly within educational contexts. First, unequal access to technology can deepen the digital divide. Not all students have reliable internet connections and/or adequate devices and hardware [51], nor the digital literacy required to use these tools effectively [52]. Rather than reducing inequalities, the use of LLMs may exacerbate them—especially if access to more advanced models is limited. For instance, while ChatGPT 3.5 is free, more sophisticated versions are subscription-based [53]. Allowing the use of LLMs for study purposes without ensuring equitable access introduces unfair advantages in assessment contexts [12].

The use of these technologies may also undermine the value of meaningful effort [4]. From perspectives related to ME theory—particularly virtue ethics—right behavior is not only about fair outcomes but also about the moral value of effort and the pursuit of excellence. Moral behavior involves not choosing the easiest path, but the one that fosters personal growth and intellectual integrity, which is often not the most convenient or effortless route [48]. Individual engagement in the learning process and commitment to personal development are integral to ethical learning [54]. For example, summarizing a text to study it for an exam after a careful and reflective reading can be considered more virtuous, and educationally enriching, than relying on an LLM-generated summary that reduces engagement with the original material.

When a student uses an LLM as a substitute for original thinking, it weakens fundamental ethical principles such as authentic personal growth and the cultivation of critical reasoning—principles that universities are expected to nurture [55]. This reduced engagement also raises concerns among both educators and students—future professionals—about potential job displacement caused by excessive reliance on LLMs [56].

From this perspective, the following hypotheses are proposed:

Hypothesis 2a (H2a): A positive perception of the use of LLMs for exam preparation from the standpoint of moral equity is a necessary condition for the intention to use.

Hypothesis 2b (H2b): A positive perception of the use of LLMs for exam preparation from the standpoint of moral equity is positively associated with the intention to use.

2.3. Consequentialism

Consequentialism (CO) holds that an action is morally valid if it produces a net balance of positive outcomes [57], whether from a selfish standpoint—focused on personal benefit [58]—or a utilitarian one, oriented toward collective well-being [59].

From an egoistic perspective, the use of LLMs for study purposes can be considered morally desirable since it facilitates learning, improves performance, and accelerates academic progress [39]. As long as no explicit rules are violated and no harm is caused to others, personal benefit may even be viewed as a moral imperative [60]. The justification for using LLMs from a consequentialist standpoint can also be reinforced by hedonic motivations [53].

However, under the same logic, LLM use may become counterproductive if it negatively affects learning or performance due to errors, biases, or "hallucinations" generated by the model [61]. Moreover, if the use of LLMs is socially disapproved of in certain institutions, its use could harm a student's reputation [55,62].

From a utilitarian perspective, significant benefits can also be argued in favor of using LLMs, given their potential to enhance collective well-being. These tools can improve understanding, reduce anxiety, and increase academic confidence when preparing exams by providing immediate feedback [38]. When used effectively, they can relieve students of tedious academic tasks and support the development of creativity and critical thinking [55]. Additionally, they free up time for other valuable activities such as rest, social engagement, or volunteering—goals increasingly valued in higher education [63]. Improving student well-being is a traditional consequentialist objective [64] and contributes to the overall improvement of educational outcomes [65].

Nevertheless, consequentialist counterarguments also exist. Uncritical use of LLMs may lead to dependency, reduce autonomous learning, and standardize language—thus limiting creativity and cultural diversity [8]. Furthermore, unequal access—due to technological or linguistic barriers—may exacerbate educational inequality [66], and the high energy consumption of these systems [67] raises environmental concerns from a global utilitarian ethics perspective.

Based on this, the following hypotheses are proposed:

Hypothesis 3a. (H3a): A positive perception of the use of LLMs for exam preparation from the standpoint of consequentialism is a necessary condition for the intention to use.

Hypothesis 3b. (H3b): A positive perception of the use of LLMs for exam preparation from the standpoint of consequentialism is positively associated with the intention to use.

2.4. Deontology

Deontology (DE), especially in its contractualist interpretation, occupies a central role by maintaining that moral evaluation should stem from compliance with universally endorsed norms rather than from the mere consequences of one's actions [68]. From this standpoint, individuals are perceived as operating within an implicit social contract that defines mutual moral duties and expectations between themselves and the broader community [22].

This perspective introduces the notion of moral duties that bind individuals beyond the consequences of their actions [68] and emphasizes the importance of obligations, particularly as students are citizens and, in many cases, future professionals in positions of responsibility [69]. In the university context, especially when part or all of the educational costs are not directly borne by students, it is reasonable to assume the existence of an implicit duty toward those who finance their studies. This duty involves striving to achieve the established learning objectives, which in most cases entails successfully passing examinations. Therefore, from this perspective, the use of LLMs as a study tool may be considered morally justifiable when applied honestly, transparently, and for legitimate educational purposes [2] such as preparing exams.

When such tools are at least partially free, as is the case for many LLMs, or when their use is institutionally encouraged—a practice that is increasingly common among educators [6,8]—they rest on an equitable foundation that reinforces their moral legitimacy. From a DE perspective, using LLMs thoughtfully and critically promotes students' intellectual autonomy, strengthens their capacity for independent thinking, and may support their moral duty to optimize their educational development [55].

Nevertheless, the use of LLMs can become morally problematic from a deontological standpoint when it violates collectively agreed-upon rules. If their use is explicitly prohibited or restricted by an educational institution [55], employing these tools to prepare for exams would contravene a shared agreement, thereby invalidating their moral justification. Likewise, if LLM use limits students' exposure to diverse approaches or perspectives—essential for deep learning [12]—it would breach their duty to maximize the depth of their education.

Additionally, if some students are unaware of LLM capabilities or lack the resources to use them, it creates unequal and non-consensual advantages [70], which can be deemed unjust under the equity principles proposed by DE. Finally, if we accept that the natural environment is also part of the implicit contract between individuals and society, students should take into account the high carbon footprint associated with LLM use. Therefore, we propose the following:

Hypothesis 4a. (H4a): A positive perception of the use of LLMs for exam preparation from the standpoint of deontology is a necessary condition for the intention to use.

Hypothesis 4b. (H4b): A positive perception of the use of LLMs for exam preparation from the standpoint of deontology is positively associated with the intention to use.

3. Materials and methods

3.1. Population and sampling

The target population of this study consisted of undergraduate students in the social sciences—particularly in fields such as economics, business management, and social work—at two Spanish universities. This profile is appropriate given that these programs combine competencies in written and oral communication, analytical reasoning, and a humanistic understanding of social dynamics [71]. Moreover, in Spanish universities, these degrees tend to rely heavily on final examinations as the main form of assessment [15]. Moreover, a review of the course syllabi for the programs in which the survey was administered confirmed that examinations, regardless of their specific format, constituted the common form of assessment. This was verified in 100 % of the first-year courses, that is, those that students had either completed or were currently enrolled in, indicating that the statements made by Panadero et al. [14] remain fully valid in the current context.

The sample was constructed using a combination of purposive sampling and snowball sampling. Participation was voluntary, anonymous, and self-administered through an online survey. Students were invited to share the survey link with peers in similar academic programs if they wished.

The voluntary and anonymous nature of participation encouraged honesty in responses and helped reduce potential social desirability bias by fostering the involvement of individuals with more reflective and altruistic motivations.

3.2. Data collection and sample profile

The questionnaire was administered during April, May, and June of 2025. It was written in Spanish. Respondents were provided with a Google Forms link that allowed only one response per IP address. The questionnaire required participants to answer all items related to the variables included in Fig. 1, although it did not require completion of sociodemographic questions such as gender, age, etc. The final number of valid responses analyzed was 151.

Table 1 presents the demographic profile of the survey participants. Table 1 summarizes the sociodemographic characteristics of the participants ($N = 151$). The sample consisted of 108 females (72.5 %) and 43 males (28.5 %). Most students were aged 22 years or younger (72.8 %), with 26.5 % aged 23 or older. Regarding perceived academic performance, 107 students (70.9 %) reported average or below-average results, and 44 (29.1 %) reported above-average performance. In terms of employment, 58 participants (38.4 %) had a full-time job, 38 (25.2 %) held temporary positions, and 55 (36.4 %) reported no

Table 1
Sociodemographic profile of the sample.

Variable	Proportion
Gender	Male 43 (28.5 %); Female 108 (72.5 %)
Age	≤20 years = 50 (33.11 %); 21 years = 33 (21.85 %); 22 years = 27 (17.88 %); ≥23 years= 40 (26.49 %); Nonanswered= 1 (0.66 %)
Perceived academic performance	On the average or below 107 (70.86 %); Above the average 44 (29.14 %)
Complementary job	Full time job=58 (38.41 %); Temporal job 38 (25.17 %); No job 55 (36.42 %)

employment.

Regarding the adequacy of the sample size, this depends on the analytical techniques employed. For NCA, although no strict minimum sample size is required, assuming a uniform distribution for the ceiling-envelopment-free disposal hull, the statistical power approaches 100 % for detecting effect sizes of at least 0.1, which are considered empirically meaningful [72].

The adequacy of the sample for regression analysis was evaluated in terms of statistical power using G*Power 3.1.0 [73]. With a sample of 151 observations, the analysis yields a statistical power of 0.80 (80 %) to detect a small-to-moderate effect size ($f^2 = 0.075$) at a 5 % significance level, assuming three independent variables. This corresponds to a minimum detectable R^2 of 6.98 %, which is substantially lower than the effect sizes reported in prior studies referenced in this research. To assess the significance of path coefficients, the size of the sample provides an 80 % power to detect small effects ($f^2 = 0.05$) at the 5 % significance level.

3.3. Questionnaire and measurement model

The questionnaire began with the following prompt: "Imagine that you are studying for a supervised exam in which you will only be allowed to use a basic scientific calculator if calculations are required. Now, think about using an AI tool based on a Large Language Model, such as ChatGPT, to prepare the exam. It will explain difficult concepts, prepare summaries, generate multiple-choice questions, or help you organize the study schedule.

Please indicate the extent to which you agree with the following statements related to this scenario, using a scale from 0 to 10, where 0 means 'strongly disagree' and 10 means 'strongly agree'."

The items used to measure the variables represented in Fig. 1 are detailed in Table 2. The scale used to assess the dependent variable—behavioral intention to use LLMs in the given scenario—was adapted from Venkatesh and Davis [45]. The scales for the ethical dimensions (ME, CO and DE) were adapted from the MES proposed by Shawver and Sennetti [32] similarly to Pelegrín-Borondo et al. [23] and Arias-Oliva et al. [74]. It is worth noting that the consequentialism dimension was constructed by combining the MES subscales corresponding to egoism and utilitarianism, following the structure proposed in Shawver and Sennetti [32].

The questionnaire was initially shared with four professors and two students to gather feedback on its readability and ease of completion. Insights gained from this pilot test helped improve the clarity of the questionnaire, although no substantial modifications were required.

Table 2
Items, descriptive statistics, and internal reliability measures of the scales.

Items	Mean	SD	Factor loading
Actual usage (USE)			
USE= *I* use LLPs to prepare exams	6.46	3.26	1.00
Intention to use (IU) (CA=0.96, CR=0.98, AVE=0.97)			
IU1= *I* will try to use LLMs to prepare exams	6.58	3.09	0.98
IU2= *I* predict that I will use LLMs to prepare exams	6.91	2.80	0.98
Moral equity (ME) (CA=0.89, CR=0.93, AVE=0.87)			
ME1= To use LLMs to prepare exams is unjust/just	7.78	2.37	0.93
ME2= To use LLMs to prepare exams is unfair/fair	7.85	2.24	0.86
ME3= To use LLMs to prepare exams is wrong/ right	7.50	2.48	0.94
Consequentialism (CO) (CA=0.86, CR=0.91, AVE=0.76)			
CO2= LLMs to prepare exams are rewarding	6.89	2.74	0.90
CO3= LLMs to prepare exams are useless/useful	7.86	2.29	0.80
CO4= LLMs to prepare exams provide a good balance cost/benefit	7.40	2.42	0.94
Deontology (DE) (CA=0.90, CR=0.95, AVE=0.91)			
DE1= Using LLMs to prepare for an exam breaks/ respects an implicit contract with my environment.	6.74	2.52	0.95
DE2=Using LLMs to polish assignments breaks / respects what is expected for me as student	6.65	2.76	0.95

Note: SD=Standard deviation, CA=Cronbach's alpha, CR=convergent reliability and AVE=average variance extracted.

3.4. Data analysis

The quantitative methods used to evaluate the proposed model follow a sequential application of two complementary techniques: NCA and PLS-SEM. This methodological combination enables a structured approach to address the two research objectives (RO1 and RO2).

RO1 focuses on determining whether the ethical perceptions assessed are necessary conditions for the acceptance of Large Language Models (LLMs) in exam preparation. This involves testing hypotheses H1a–H4a using NCA [34,36], which allows verification of, for example, whether "a minimum perception of ME is necessary to generate IU LLMs."

RO2, in turn, tests hypotheses H1b–H4b, which examine whether certain ethical perceptions are sufficient conditions to generate intention to use (IU), and whether IU itself is sufficient to explain actual use (USE) of LLMs. This analysis is carried out using PLS-SEM, a technique suitable for small samples [75] and particularly useful for evaluating the predictive power of the proposed model [35].

NCA was implemented using the R package NCA() [76], while scale validation and PLS-SEM were conducted using the cSEM package [77]. The overall analytical procedure, integrating both techniques, follows the general protocol outlined by Richter et al. [35], with one key distinction. In this study, NCA was applied prior to the PLS-SEM analysis.

The procedure followed three steps:

Step 1. involved evaluating the internal consistency and discriminant validity of the constructs [78]. In this step, standardized factor loadings were used as construct scores.

Step 2. addressed the RO1 hypotheses using NCA. The data were first cleaned by removing outliers, defined as cases in which the difference between behavioral intention and the explanatory factor exceeded three times the standard deviation of that difference across the sample. Using scatterplots, ceiling envelopment–free disposal hull (CE-FDH) envelopes were identified, and the necessity effect size (d) was calculated for each pair of dependent and explanatory variables: USE and IU; IU and ME; IU and CO; and IU and DE [34].

In the CE-FDH method, the upper frontier of the data represents the maximum achievable level of the outcome variable (USE or IU) as a function of the explanatory variables (IU in the case of USE, or ME, CO, and DE for IU). This stepwise "envelope" delineates the empty space where no observations exist, indicating that certain levels of the outcome variable are impossible without a minimum level of the explanatory variable. The area of this empty space determines the effect size (d), which quantifies the degree of necessity of the condition [34].

Hypotheses H1a–H4a were considered supported when $d > 0.1$ and, according to the permutation test, $p < 0.05$ [36]. Additionally, bottleneck tables were constructed to identify the minimum level of each input variable required to achieve a specific level of the outcome variable.

Step 3. addressed RO2 by estimating the model presented in Fig. 1 with PLS-SEM. The model was fitted through percentile bootstrapping with 10,000 subsamples. This step evaluated the econometric robustness of the proposed model—examining both model fit and predictive capacity—and tested the sufficiency hypotheses (H1b–H4b). A hypothesis was considered supported when the corresponding path coefficient exhibited a p-value below 0.05.

4. Results

4.1. Descriptive statistics and assessment of measurement model

The mean scores and standard deviations of the items included in the study are presented in Table 2. It is important to note that, since an 11-point Likert scale (ranging from 0 to 10) was used, a score of 5 represents a neutral or indifferent evaluation.

The prevailing behavioral intention toward the use of LLMs as study material is favorable, with the average of its two items ranging between 6 and 7. A similar evaluation is observed for the effective use of LLMs, with its mean score being only slightly lower. The assessment of items related to ethical judgments is even more positive than the intention to use: items related to ME and CO have mean scores between 7 and 8, while those related to DE are slightly below 7.

Table 2 also shows that the scales used exhibit adequate internal reliability. In almost all cases, both Cronbach's alpha and composite reliability exceed the recommended threshold of 0.70; the factor loadings of the items are above 0.70, and the average variance extracted (AVE) is greater than 0.50. Furthermore, the results in Table 3 support the discriminant validity of the scales according to the Fornell-Larcker criterion, as the square root of the AVE is greater than the Pearson correlations between the dimensions.

4.2. Results of the necessary condition analysis (research objective 1)

The graphical representation of the scatterplot and the ceiling envelopment–free disposal hull CE-FDH for the relationship between IU and USE is presented in Fig. 2. Table 4 shows the extent to which IU is a necessary condition for actual use to occur. It reveals that the degree of necessity (d) is 0.325 ($p < 0.001$), indicating a strong and significant effect; thus, hypothesis H1a is supported.

The CE-FDH for the relationship between ethical perceptions and IU is also shown in Fig. 2. Table 4 indicates that all these variables have a statistically significant status as necessary conditions for IU. The strongest necessary condition—stronger even than IU for USE—is ME, with a d value of 0.338 and $p < 0.001$, classifying it as strong and significant. The second most influential necessary condition is CO, with $d = 0.274$ (moderate-to-high) and $p < 0.001$. The effect of DE, although more modest ($d = 0.207$), is also significant ($p < 0.001$). In all cases, the values exceed the 0.1 threshold, indicating empirical relevance [34] and nearly 100 % statistical power [72]. Therefore, hypotheses H1a, H2a, H3a, and H4a are supported.

Table 5 presents the bottleneck tables for the relationship between IU and USE, as well as for ME, CO, and DE with IU. These tables identify the minimum value that each explanatory variable must reach in order for a given level of USE or IU to occur.

In the case of USE, achieving a medium level (50th percentile) requires, at minimum, reaching the 20th percentile in IU. For high levels of use (80th percentile and above), the bottleneck indicates that it is necessary to exceed the 60th percentile in IU. Finally, full use (100th percentile) requires reaching at least the 70th percentile in intention to use.

Regarding IU, reaching a neutral level (50th percentile) requires at least the 23.8th percentile in ME and the 20th percentile in DE. At this level, CO is not a necessary condition. For higher levels of IU (70th and 80th percentiles), the most demanding thresholds correspond to ME and CO, which must be at least in the 36.6th and 48th percentiles, respectively. DE is also required, though with more moderate demands—again, the 20th percentile. Finally, full acceptance of LLMs (100th percentile) requires reaching the 77.2nd percentile in ME, the 55.3rd percentile in DE, and exceeding the 70th percentile in CO.

Table 3
Correlations and discriminant validity measures (Fornell-Larcker criterion).

	USE	IU	ME	CO	DE
USE	1.000				
IU	0.905	0.983			
ME	0.567	0.654	0.909		
CO	0.630	0.700	0.827	0.881	
DE	0.608	0.690	0.779	0.793	0.953

Note: In the principal diagonal comes the squared average variance extracted and under it, Pearson's correlations.

4.3. Results of PLS-SEM analysis (research objective 2)

Table 6 presents the results of the PLS-SEM estimation, including the path coefficients (β) and their significance, measured through p-values. The influence of IU on USE is highly significant ($\beta = 0.905$, $p < 0.001$), indicating that IU functions as a sufficient condition for actual use. The effect size ($f^2 = 4.505$) is very large.

Moreover, both CO ($\beta = 0.350$, $p = 0.038$, $f^2 = 0.070$) and DE ($\beta = 0.350$, $p = 0.013$, $f^2 = 0.072$) exhibit significant path coefficients on IU, with small-to-moderate effect sizes. Therefore, both ethical perceptions can also be considered sufficient conditions for IU LLMs. In contrast, the influence of ME on IU is not statistically significant.

Regarding model fit, as shown in Table 6 and following the criteria proposed by Hair et al. [78], the quality of fit for USE is substantial ($R^2 = 81.8$ %), while for IU it is moderate ($R^2 = 54.3$ %). In both cases, the R^2 values clearly exceed the minimum threshold of 7 % established in Section 3.2. In fact, these coefficients of determination indicate a statistical power close to 100 %. Additionally, the VIF values for all variables are well below 5, indicating no serious multicollinearity issues. As for predictive relevance, it is moderate for USE ($Q^2 = 36.9$ %) and high for the IU model ($Q^2 = 50.5$ %), according to the criteria by Hair et al. [78].

5. Discussion

5.1. General discussion of results

5.1.1. Overview of the research

The findings of this study provide a broad and nuanced perspective on how ethical perceptions influence the acceptance and use of large language models (LLMs) by university students in the context of supervised exam preparation. Using a dual methodological approach—combining Necessary Condition Analysis (NCA) and Partial Least Squares Structural Equation Modeling (PLS-SEM)—the study addresses two research objectives (ROs).

The first (RO1) aimed to determine whether a minimum threshold exists in three ethical dimensions for the formation of intention to use LLMs. The second (RO2) sought to assess whether higher ethical evaluations lead to greater levels of acceptance. Ethical perceptions were measured using a modified version of the Multimensional Ethics Scale (MES) scale by Reidenbach and Robin [22], adapted by Shawver and Sennetti [32].

5.1.2. Ethical judgments as a necessary and sufficient conditions

To address RO1, the study applied NCA, which revealed that all assessed ethical dimensions—moral equity (ME), consequentialism (CO), and deontology (DE)—constitute necessary conditions for the existence of intention to use (IU) LLMs. This means that, regardless of their magnitude of effect, students must reach at least a minimum threshold in each of these ethical dimensions to even consider using LLMs as part of their study strategies. Although the MES scale captures normative frameworks that are often philosophically opposed, the results suggest that all these dimensions are relevant in ethical decision-making within this context. For instance, the fact that ME and DE are necessary conditions implies that students will only consider using LLMs if they do not perceive significant injustice or a violation of implicit contracts with educational institutions and society. Similarly, the

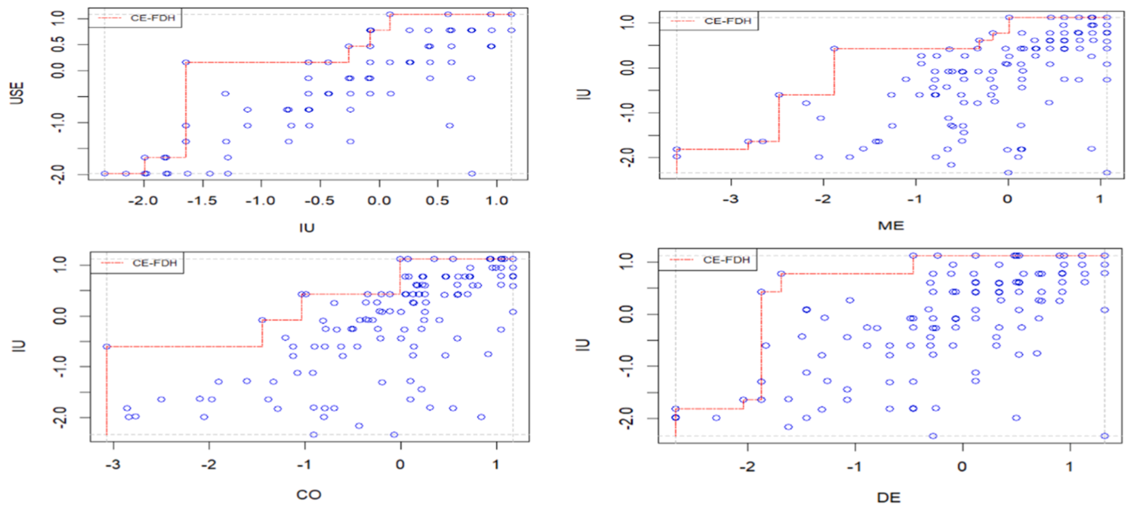

Fig. 2. Ceiling envelopment–free disposal hull depicting the relationship between intention to use and actual use, and the relationship between moral equity, consequentialism, and deontology with intention to use. Source: Authors' own work by using Richter et al. [35] and Dul et al. [36].

Table 4
Effect Sizes for the necessity status of the assessed explanatory variables.

Relation	Effect size	p-value	Decision on hypothesis
IU→USE	0.325	<0.001	H1a=Supported
ME→IU	0.338	<0.001	H2a=Supported
CO→IU	0.274	<0.001	H3a=Supported
DE→IU	0.207	<0.001	H4a=Supported

Table 5
Bottleneck tables for USE and IU.

Output (USE)	IU	Output (IU)	ME	CO	DE
0	NN	0	NN	NN	NN
10	10	10	NN	NN	NN
20	20	20	23.8	NN	20
30	20	30	23.8	NN	20
40	20	40	23.8	NN	20
50	20	50	23.8	NN	20
60	20	60	36.6	38.3	20
70	20	70	36.6	48.0	20
80	60	80	36.6	48.0	20
90	70	90	73.5	72.3	24.7
100	70	100	77.2	72.3	55.3

Note: (a) Quantities come in percentage (b) The output for ME, RE, CO and DE is IU; and the output of IU is USE, (c) NN stands for no necessary.

necessity of CO indicates that students are only likely to consider using these tools if they anticipate positive consequences from egoistic and utilitarian perspectives. This finding aligns with recent studies highlighting the growing ethical awareness among university students, such as the one reported by Barbosa [79] in the Portuguese context.

Regarding RO2, the PLS-SEM results reveal that both consequentialism and deontology exert statistically significant and sufficient effects on intention to use. In contrast, moral equity does not have a statistically significant effect. This divergence between necessity and sufficiency is theoretically important, as it shows that a condition may be necessary without being sufficient. In other words, while a perception of moral equity is indispensable for IU to exist, it is not enough by itself to explain it. In fact, ME has the largest effect size as a necessary condition, indicating that no student reports a high level of intention to use LLMs without perceiving a high degree of fairness. However, once the minimum threshold for ME is met, further increases in perceived fairness do not lead to greater IU. In contrast, consequentialism and deontology exhibit lower necessity thresholds. Once those thresholds are exceeded, a relative lack in one of the dimensions can be offset by the presence of the other to promote higher levels of acceptance. Moral equity, however, does not operate in a compensatory fashion—its influence stabilizes once its threshold is met.

It also must be remarked as a key finding that intention to use is confirmed as both a necessary and sufficient condition for explaining actual use (USE) of LLMs.

Results in RO1 and RO2 justifies the need to combine complementary methodologies to understand complex phenomena [33,35] such as how ethics influences decision making. From a decision-making perspective, this means that for a student to even consider using LLMs over alternative tools, they must first perceive such use as fair, deontologically acceptable, and offering some utility. Only under these necessary conditions does IU become possible. Within that context, a greater perception that LLMs help fulfill obligations toward society, their immediate environment, and their academic institution—as well as being effective tools for academic achievement—translates into a higher intention to use.

This results in RO1 and RO2 can be also understood through the conceptual difference between acceptability and acceptance of technology [80]. Moral equity primarily determines acceptability, that is, whether the use of LLMs is perceived as fair, legitimate, and ethically

Table 6
Results of the PLS-SEM estimation of the model developed in Section 2.

Path	β	SD	f^2	VIF	Students' t	p-value	Decision
IU→USE	0.905	0.030	4.505	1	30.249	<0.001	H1b: Supportedt
ME→IU	0.108	0.132	0.007	3.64	0.818	0.414	H2b: Not supported
CO→IU	0.350	0.169	0.070	3.854	2.078	0.038	H4b: Supported
DE→IU	0.329	0.133	0.077	3.088	2.48	0.013	H5b: Supported

Notes: (a) β stands por path coefficient and SD standard deviation (b) The determination coefficient for USE is 81.8 % and for IU is 54.3 %. (b) El Q^2 de Stone-Geisser es 39.6 % para USE and 50.5 % para IU.

permissible. From a virtue ethics perspective, this dimension reflects the moral threshold that must be met before any behavioral intention can emerge. However, acceptance and actual use require additional motivational drivers. Once students perceive the technology as morally acceptable, the degree of their intention to use LLMs depends mainly on consequentialist and deontological considerations. The former relates to perceived efficiency gains—such as completing tasks in less time or improving academic outcomes—while the latter reflects a sense of duty to meet academic expectations and learning goals assumed toward instructors, peers, and family. Consequently, moral equity establishes the ethical legitimacy of use, whereas consequentialism and deontology provide the motivational force that translates acceptability into actual acceptance and use.

5.1.3. Contextualizing the results of this paper in the literature

The fact that consequentialism—often associated with rational egoism and utilitarianism—emerges as the most statistically significant explanatory dimension supports findings from Biloš and Budimir [53], Boubker [1], Dahri et al. [81] Mustofa et al. [82], Sobaih et al. [83] and Strzelecki [84]. These studies, grounded in technology acceptance models such as TAM and UTAUT, highlight the relevance of consequentialist constructs like perceived usefulness, ease of use, and hedonic motivation in the adoption of LLMs in academic settings. It is also notable that using the same MES scale from the faculty's perspective, Pelegrín-Borondo et al. [27] found that moral egoism significantly influences faculty willingness to promote LLM use among students.

The fact that IU is a necessary condition and highly statistical significant to explain USE, result validates the core assumptions of technology acceptance models such as TAM [30] and UTAUT [31], and aligns with findings from Sobaih et al. [83] and Strzelecki [84] in students' use of LLMs.

Overall, ethical perceptions of LLM use are positive. Mean scores are above the midpoint of the scale, especially for moral equity and consequentialism, suggesting that students tend to view these tools as both fair and useful—both in terms of justice and outcomes. This favorable view may partly stem from the perception of LLMs as democratizing tools that provide personalized access to knowledge regardless of socioeconomic background [49]. This interpretation is consistent with Shahzad et al. [85], who also report a generally positive valuation of AI use in educational settings in China.

Moreover, the statistical significance of the deontological dimension is consistent with prior literature. This ethical framework has been shown to inhibit academically questionable behaviors such as plagiarism and cheating [28,29], as well as to explain faculty members' intention to use LLMs in their teaching [27]. This finding can also be interpreted through the lens of technology acceptance models, which have shown that perceived social norms favoring the use of LLMs in academia are positively associated with their adoption [83,84,86].

5.2. Theoretical implications

This study makes a significant contribution to the literature on technology acceptance in educational settings by incorporating a broader ethical perspective alongside a robust methodological approach. While previous research has primarily focused on factors such as perceived usefulness, ease of use, or social influence [31], this study introduces normative elements drawn from moral philosophy to explain the adoption of emerging technologies. The integration of the MES scale [22] into a technology acceptance model provides a more comprehensive conceptual framework, enabling an analysis not only of the functionality of technology but also of its ethical legitimacy from multiple normative perspectives.

Moreover, the study demonstrates the value of combining NCA and PLS-SEM to evaluate the influence of ethical perception in decision making. This complementary methodological approach allows for the identification of both necessary and sufficient conditions, overcoming the limitations of traditional statistical models, which tend to focus exclusively on sufficient causal relationships and often overlook the minimum thresholds required for certain behaviors to occur [34]. The finding that moral equity is a necessary but not sufficient condition highlights the complexity with which ethical judgments influence decision-making. This complexity is further underscored by the observation that moral equity has the largest effect size as a necessary condition, while showing the least statistical relevance in the sufficiency model.

Likewise, the finding that all evaluated ethical dimensions are necessary conditions for the formation of intention to use—and that two of them, consequentialism and deontology, also act as sufficient conditions—is consistent with the original proposition of the MES scale, which recognizes that ethical decisions are based on the coexistence of multiple normative frameworks [22].

Empirical evidence shows that consequentialism and deontology are statistically significant in the PLS-SEM model. This suggests that among students who perceive LLM use as surpassing minimum ethical thresholds, a greater perception of favorable consequences translates into a higher level of acceptance. Such perceptions may be motivated by the expectation of achieving better academic results—or the same results with less effort. This enhanced performance can also be interpreted through a deontological lens, insofar as many students view academic success as part of an implicit contract with their families and with society. In this sense, the use of LLMs may be seen as an adaptive response to the Spanish university context, where academic performance plays a central role in evaluating student achievement [14].

The necessary nature of the moral equity dimension suggests that students' decisions are not based solely on utilitarian or deontological reasoning. Rather, their judgments are shaped by a more complex ethical structure that should be recognized and incorporated into theoretical models applied to the study of technology acceptance. In this regard, studies such as Shawver and Sennetti [32] provide valuable conceptual tools for the teaching and analysis of ethical decision-making in educational contexts, like those addressed in this study.

5.3. Practical implications

The findings of this study have meaningful practical implications for education policymakers, university administrators, and instructors. First and foremost, the fact that the acceptance and use of LLMs are shaped by ethical judgments suggests that their integration into academic environments must go beyond mere technical access. It is essential to accompany their adoption with ethical training that equips students to make informed and responsible decisions [24]. Institutions could design explicit ethical rubrics and guidelines defining acceptable uses of LLMs for exam preparation—such as question generation, self-testing, and conceptual explanation—thus operationalizing deontological principles into clear behavioral standards.

The finding that consequentialism and deontology function as both necessary and sufficient conditions for intention to use provides a clear pathway for promoting responsible adoption. Institutional campaigns could focus on highlighting tangible benefits—such as improved academic performance, time savings, or support for autonomous learning—while also raising awareness of potential risks, including technological dependency or the erosion of cognitive skills [87]. The relevance of deontology further suggests that clear guidelines and codes of conduct regarding LLM use can facilitate responsible behavior.

The fact that all three ethical dimensions are necessary conditions implies that their absence can hinder IU even if the functional benefits are evident. This underscores the importance of inclusive policies that ensure equitable and legitimate access to LLMs [9]. For instance, students who perceive LLMs as unfair or inconsistent with educational equity may be reluctant to use them, despite acknowledging their effectiveness. Universities should provide institutional access to reliable LLM platforms with comparable functionality to paid versions. This would prevent economic disparities among students from translating into unequal learning conditions and tools to exam preparation, thereby reinforcing the principle of moral equity identified in this study.

Given that intention to use is both a necessary and sufficient condition for actual use, institutions should create environments that foster positive attitudes toward these tools [56]. This can be achieved through digital literacy programs, training in the ethical use of LLMs [87], thematic seminars, or collaborative activities that allow students to explore their potential in guided settings [85]. Higher education institutions could implement supervised AI-assisted exam simulations, where students practice reasoning and problem-solving with LLMs under instructor guidance. Such sessions would promote reflective learning and reduce misuse by emphasizing critical evaluation of AI-generated responses. Likewise, it may be incorporated short digital literacy modules focused on self-regulation and critical thinking when using LLMs for exam study. Embedding these modules in first-year courses would help students build responsible and autonomous learning habits from the outset.

From a pedagogical standpoint, these results invite educators to rethink assessment methods. If students perceive that LLMs enhance their learning, they are likely to use them—even in the absence of explicit regulation. Therefore, rather than relying on prohibitions, institutions should consider process-oriented assessment systems that encourage creativity, critical reflection, and the integration of multiple sources [88].

6. Conclusions

6.1. Principal findings

This study provides empirical evidence on the role that distinct ethical dimensions play in the acceptance and use of large language models (LLMs) by university students during their preparation for supervised exams. Through a dual methodological approach—combining Necessary Condition Analysis (NCA) and Partial Least Squares Structural Equation Modeling (PLS-SEM)—the results enhance our understanding of the phenomenon from both normative and predictive perspectives.

First, it was confirmed that all three ethical dimensions assessed—moral equity, consequentialism, and deontology—function as necessary conditions for the formation of intention to use (IU). This implies that unless students reach a minimum threshold in each of these dimensions, it is unlikely they will develop a favorable attitude toward the use of LLMs.

Furthermore, ME, CO, and DE also showed statistically significant and sufficient effects in the PLS-SEM model, suggesting that once minimum ethical thresholds are surpassed, the acceptance of LLMs increases only when students' perceptions in these ethical domains are also high. In contrast, relativism did not exhibit a significant sufficient effect, indicating that while its presence may be necessary, it does not directly drive a higher level of acceptance.

Additionally, the analysis confirmed that intention to use is both a necessary and sufficient condition for explaining the actual use (USE) of LLMs. This finding reinforces the central role of IU in classic technology adoption models such as TAM [30] and UTAUT [31], even when ethical variables are integrated into the analysis.

Overall, the study advances theoretical, empirical, and practical understanding of how ethical considerations influence student behavior regarding emerging educational technologies, particularly in contexts where academic integrity and responsible use are paramount.

6.2. Limitation of the study

Despite the methodological rigor of the approach employed, this study presents several limitations that should be considered when interpreting the results.

First, the sample consists exclusively of university students from social science–related programs, which restricts the generalizability of the findings to other educational domains, such as STEM disciplines. However, this focus also provides a diverse academic landscape in itself, encompassing both quantitative, logic-oriented subjects (e.g., statistics, accounting) and humanistic ones (e.g., social work, sociology), where the forms of examination vary considerably. This heterogeneity strengthens the relevance of the results for understanding exam preparation practices in contexts that combine different cognitive and evaluative demands.

Additionally, the data were collected through self-reported measures, which may introduce social desirability bias—particularly in responses related to ethical judgments. As noted by Ma et al. [89], individuals with higher levels of education tend to exhibit a greater willingness to adopt artificial intelligence, which may influence how students respond to questions about emerging technologies.

The study was conducted solely with Spanish participants, which allows for the extrapolation of findings to culturally similar contexts, but limits their applicability to culturally distinct environments. Previous cross-cultural research has shown that the influence of explanatory variables on the intention to use of generative language models can vary significantly across cultural settings [81,90].

Another important limitation is the cross-sectional nature of the study. Since perceptions and behaviors were measured at a single point in time, it is not possible to draw conclusions about the evolution of ethical attitudes or changes in LLM use over time.

Finally, the proposed model focuses exclusively on ethical variables as antecedents of intention to use, omitting other potentially relevant factors such as academic performance, institutional pressure, or levels of digital literacy.

6.3. Further research

The limitations of this study serve as a motivation for future research. One possibility is to expand the proposed model by incorporating contextual or institutional variables, such as the presence of academic ethics codes, institutional regulations governing AI use, or the degree of instructor oversight. This would enable an analysis of how individual ethical perceptions interact with broader normative and organizational environments, offering a more comprehensive understanding of the conditions that shape responsible technology adoption.

It would also be valuable to replicate the study across different populations and educational levels, including other countries and non-university contexts. Such efforts would allow for cross-cultural comparisons in the ethics of technology use and help assess the universality of the factors identified as necessary or sufficient for acceptance.

Finally, longitudinal studies could explore how ethical perceptions and the use of LLMs evolve over time as these technologies become more deeply integrated into educational processes. This approach would facilitate the identification of dynamic changes, potential habituation effects, or shifts in academic morality associated with the prolonged use of generative artificial intelligence.

Funding

This research was supported by Telefonica and the Telefonica Chair on Smart Cities of the Universitat Rovira i Virgili and Universitat de Barcelona (project number: 42. DB.00.18.00).

Institutional review board statement

(1) All participants received detailed written information about the study and procedure; (2) no data directly or indirectly related to the health of the subjects were collected, and therefore the Declaration of Helsinki was not mentioned when informing the subjects; (3) the anonymity of the collected data was ensured at all times; (4) the research received a favorable evaluation from the Ethics Committee of a

researchers' institution (CE_20,250,710_10_SOC).

Informed consent statement

All respondents gave permission for the processing of their responses for the content of this publication.

Data availability statement

The data supporting the analysis is available at by demanding to any author.

CRediT authorship contribution statement

Antonio Pérez-Portabella: Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Mario Arias-Oliva:** Validation, Project administration, Investigation, Funding acquisition, Conceptualization. **Graciela Padilla-Castillo:** Supervision, Resources, Project administration, Methodology. **Jorge de Andrés-Sánchez:** Writing – original draft, Visualization, Software, Data curation.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Mario Arias-Oliva reports article publishing charges, statistical analysis, and writing assistance were provided by Cátedra Smart Cities Universitat Rovira i Virgili. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Boubker O. From chatting to self-educating: can AI tools boost student learning outcomes? Expert Syst Appl 2024;238:121820. https://doi.org/10.1016/j.eswa.2023.121820.

[2] Lo CK. What is the impact of ChatGPT on education? A rapid review of the literature. Educ Sci 2023;13(4):410. https://doi.org/10.3390/educsci13040410.

[3] Parker L, Carter C, Karakas A, Loper AJ, Sokkar A. Graduate instructors navigating the AI frontier: the role of ChatGPT in higher education. Comput Educ Open 2024;6:100166. https://doi.org/10.1016/j.caeo.2024.100166.

[4] Pahi K, Hawlader S, Hicks E, Zaman A, Phan V. Enhancing active learning through collaboration between human teachers and generative AI. Comput Educ Open 2024;6:100183. https://doi.org/10.1016/j.caeo.2024.100183.

[5] Tzirides AO(Olnancy), Zapata G, Kastania NP, Saini AK, Castro V, Ismael SA, You Y, dos Santos TA, Searsmith D, O'Brien C, Cope B, Kalantzis M. Combining human and artificial intelligence for enhanced AI literacy in higher education. Comput Educ Open 2024;6:100184. https://doi.org/10.1016/j.caeo.2024.100184.

[6] Yu H. Reflection on whether Chat GPT should be banned by academia from the perspective of education and teaching. Front Psychol 2023;14:1181712. https://doi.org/10.3389/fpsyg.2023.1181712.

[7] Meyer JG, Urbanowicz RJ, Martin PC, O'Connor K, Li R, Peng PC, Moore JH. ChatGPT and large language models in academia: opportunities and challenges. BioData Min 2023;16(1):20. https://doi.org/10.1186/s13040-023-00339-9.

[8] Tlais S, Alkhatib A, Hamdan R, HajjHussein H, Hallal K, Malti WE. Artificial intelligence in higher education: early perspectives from Lebanese stem faculty. TechTrends; 2025. p. 1–9. https://doi.org/10.1007/s11528-025-01061-5.

[9] Zhou J, Müller H, Holzinger A, Chen F. Ethical ChatGPT: concerns, challenges, and commandments. Electron (Basel) 2024;13(17):3417. https://doi.org/10.3390/electronics13173417.

[10] Eke DO. ChatGPT and the rise of generative AI: threat to academic integrity? J Responsible Technol 2023;13:100060. https://doi.org/10.1016/j.jrt.2023.100060.

[11] Moorhouse BL, Yeo MA, Wan Y. Generative AI tools and assessment: guidelines of the world's top-ranking universities. Comput Educ Open 2023;5:100151. https://doi.org/10.1016/j.caeo.2023.100151.

[12] Oates A, Johnson D. ChatGPT in the classroom: evaluating its role in fostering critical evaluation skills. Int J Artif Intell Educ 2025:1–32. https://doi.org/10.1007/s40593-024-00452-8.

[13] Aler Tubella A, Mora-Cantallops M, Nieves JC. How to teach responsible AI in higher education: challenges and opportunities. Ethics Inf Technol 2023;26(1):3. https://doi.org/10.1007/s10676-023-09733-7.

[14] Panadero E, Fraile J, Fernández Ruiz J, Castilla-Estévez D, Ruiz MA. Spanish university assessment practices: examination tradition with diversity by faculty. Assess Eval High Educ 2018;44(3):379–97. https://doi.org/10.1080/02602938.2018.1512553.

[15] San Martín Gutiérrez S, Jiménez Torres N, Jerónimo Sánchez-Beato E. La evaluación del alumnado universitario en el Espacio Europeo de Educación Superior. Aula Abierta 2016;44(1):7–14. https://doi.org/10.1016/j.aula.2015.03.003.

[16] Zimmerman BJ. Becoming a self-regulated learner: an overview. Theory Pr 2002;41(2):64–70. https://doi.org/10.1207/s15430421tip4102_2.

[17] Red Española de Agencias de Calidad Universitaria, (2021). Directrices y orientaciones para la elaboración y evaluación de los planes de estudios de títulos de grado y máster universitario en diferentes modalidades de enseñanza.

[18] Moreno JM. Educación en la sombra en España: cómo las clases particulares se están convirtiendo en un bien de primera necesidad.ESADE EcPol – Center for Economic Policy. https://www.esade.edu/ecpol/es/publicaciones/educacion-en-la-sombra-en-espana-como-las-clases-particulares-se-estan-convirtiendo-en-un-bien-de-primera-necesidad/.

[19] Haleem A, Javaid M, Qadri MA, Suman R. Understanding the role of digital technologies in education: a review. Sustain Oper Comput 2022;3:275–85. https://doi.org/10.1016/j.susoc.2022.05.004.

[20] Thornberg R. A classmate in distress: schoolchildren as bystanders and their reasons for how they act. Soc Psychol Educ 2007;10:5–28. https://doi.org/10.1007/s11218-006-9009-4.

[21] Grühn D, Cheng Y. A self-correcting approach to multiple-choice exams improves students' learning. Teach Psychol 2014;41(4):335–9. https://doi.org/10.1177/0098628314549706.

[22] Reidenbach RE, Robin DP. Toward the development of a multidimensional scale for improving evaluations of business ethics. J Bus Ethics 1990;9(8):639–53. https://doi.org/10.1007/BF00383391.

[23] Pelegrín-Borondo J, Arias-Oliva M, Murata K, Souto-Romero M. Does ethical judgment determine the decision to become a cyborg? J Bus Ethics 2020;161(1):5–17. https://doi.org/10.1007/s10551-018-3970-7.

[24] Cunha M, Figueiredo J, Breia J, Pina J, Almeida S, Oliveira T. Morality and ethical acting in higher education students. European Proceedings of Social and Behavioural Sciences; 2016. p. 98–106.

[25] Jung I. Ethical judgments and behaviors: applying a multidimensional ethics scale to measuring ICT ethics of college students. Comput Educ 2009;53(3):940–9. https://doi.org/10.1016/j.compedu.2009.05.011.

[26] Leonard LN, Riemenschneider CK, Manly TS. Ethical behavioral intention in an academic setting: models and predictors. J Acad Ethics 2017;15:141–66. https://doi.org/10.1007/s10805-017-9273-2.

[27] Pelegrín-Borondo JP, Pascual CO, Pascual LB, Milon AG. Impact of ethical judgment on university professors encouraging students to use AI in academic tasks. The leading role of smart ethics in the digital world. Universidad de La Rioja; 2024. p. 53–61.

[28] Prashar P, Gupta Y, Dwivedi YK. Plagiarism awareness efforts, students' ethical judgment and behaviors: a longitudinal experiment study on ethical nuances of plagiarism in higher education. Stud High Educ 2024;49(6):929–55. https://doi.org/10.1080/03075079.2023.2253835.

[29] Yang SC. Ethical academic judgments and behaviors: applying a multidimensional ethics scale to measure the ethical academic behavior of graduate students. Ethics Behav 2012;22(4):281–96. https://doi.org/10.1080/10508422.2012.672907.

[30] Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989;13(3):319–40. https://doi.org/10.2307/249008.

[31] Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425–78. https://doi.org/10.2307/30036540.

[32] Shawver TJ, Sennetti JT. Measuring ethical sensitivity and evaluation. J Bus Ethics 2009;88(4):663–78. https://doi.org/10.1007/s10551-008-9973-z.

[33] Dul J. A different causal perspective with necessary condition analysis. J Bus Res 2024;177:114618. https://doi.org/10.1016/j.jbusres.2024.114618.

[34] Dul J. Necessary condition analysis (NCA): logic and methodology of "necessary but not sufficient" causality. Organ Res Methods 2016;19(1):10–52. https://doi.org/10.1177/1094428115584005.

[35] Richter NF, Schubring S, Hauff S, Ringle CM, Sarstedt M. When predictors of outcomes are necessary: guidelines for the combined use of PLS-SEM and NCA. Ind Manag Data Syst 2020;120(12):2243–67. https://doi.org/10.1108/IMDS-11-2019-0638.

[36] Dul J, der Laan E, Kuik R. A statistical significance test for necessary condition analysis. Organ Res Methods 2018;23(2):385–95. https://doi.org/10.1177/1094428118795272.

[37] Ajzen I. Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior. J Appl Soc Psychol 2002;32(4):665–83. https://doi.org/10.1111/j.1559-1816.2002.tb00236.x.

[38] Hasanein AM, Sobaih AEE. Drivers and consequences of ChatGPT use in higher education: key stakeholder perspectives. Eur J Investig Health Psychol Educ 2023;13(11):2599–614. https://doi.org/10.3390/ejihpe13110181.

[39] Leite H. Artificial intelligence in higher education: research notes from a longitudinal study. Technol Forecast Soc Change 2025;215:124115. https://doi.org/10.1016/j.techfore.2025.124115.

[40] Lo CK, Hew KF, Jong MS. The influence of ChatGPT on student engagement: a systematic review and future research agenda. Comput Educ 2024;219:105100. https://doi.org/10.1016/j.compedu.2024.105100.

[41] Dai Y, Lai S, Lim CP, Liu A. ChatGPT and its impact on research supervision: insights from Australian postgraduate research students. Australas J Educ Technol 2023;39(4):74–88. https://doi.org/10.14742/ajet.8843.

[42] Almassaad A, Alajlan H, Alebaikan R. Student perceptions of generative artificial intelligence: investigating utilization, benefits, and challenges in higher education. Systems 2024;12(10). https://doi.org/10.3390/systems12100385.

[43] Hemmingsen M. Act consequentialism and the gamer's dilemma. Ethics Inf Technol 2025;27(2):19. https://doi.org/10.1007/s10676-025-09827-4.

[44] Sheppard BH, Hartwick J, Warshaw PR. The theory of reasoned action: a meta-analysis of past research with recommendations for modifications and future research. J Consum Res 1988;15(3):325–43. https://doi.org/10.1086/209170.

[45] Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manage Sci 2000;46(2):186–204. https://doi.org/10.1287/mnsc.46.2.186.11926.

[46] Killen M. The origins of morality: social equality, fairness, and justice. Philos Psychol 2018;31(5):767–803. https://doi.org/10.1080/09515089.2018.1486612.

[47] Nussbaum M, Sen A. The quality of life. Clarendon Press; 1993.

[48] MacIntyre A. After virtue: a study in moral theory. 2nd ed. University of Notre Dame Press; 1984.

[49] Qadir J. Engineering education in the era of ChatGPT: promise and pitfalls of generative AI for education. In: 2023 IEEE Global Engineering Education Conference (EDUCON); 2023. p. 1–9. https://doi.org/10.1109/EDUCON54358.2023.10125121.

[50] Dwivedi YK, Kshetri N, Hughes L, Slade EL, Jeyaraj A, Kar AK, Wright R. Opinion paper: "so what if ChatGPT wrote it?" multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. Int J Inf Manage 2023;71:102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642.

[51] Toh Y, Looi CK. Transcending the dualities in digital education: a case study of Singapore. Front Digit Educ 2024;1(2):121–31. https://doi.org/10.1007/s44366-024-0002-2.

[52] Alier M, García-Peñalvo F, Camba JD. Generative artificial intelligence in education: from deceptive to disruptive. Int J Interact Multimed Artif Intell 2024;8(5):5–14. https://doi.org/10.9781/ijimai.2024.02.011.

[53] Biloš A, Budimir B. Understanding the adoption dynamics of ChatGPT among generation Z: insights from a modified UTAUT2 model. J Theor Appl Electron Commer Res 2024;19(2):863–79. https://doi.org/10.3390/jtaer19020045.

[54] Dalbert C, Schneidewind U, Saalbach A. Justice judgments concerning grading in school. Contemp Educ Psychol 2007;32(3):420–33. https://doi.org/10.1016/j.cedpsych.2006.05.003.

[55] Licht K, de F. Generative artificial intelligence in higher education: why the "banning approach" to student use is sometimes morally justified. Philos Technol 2024;37(3):113. https://doi.org/10.1007/s13347-024-00799-9.

[56] Schmidt DA, AlBloushi B, Thomas A, Magalhaes R. Integrating artificial intelligence in higher education: perceptions, challenges, and strategies for academic innovation. Comput Educ Open 2025:100274. https://doi.org/10.1016/j.caeo.2025.100274.

[57] Sinnott-Armstrong, W. (2023). Consequentialism. In E. N. Zalta (editor), The stanford encyclopedia of philosophy.

[58] Deigh J. Egoism. An introduction to ethics. Cambridge University Press; 2010. p. 25–55. https://doi.org/10.1017/CBO9780511750519.

[59] Rae S. Moral choices: an introduction to ethics. Zondervan Academic; 2018.

[60] Rand A. The virtue of selfishness. New American Library; 1964.

[61] Parviz M. AI in education: comparative perspectives from STEM and non-STEM instructors. Comput Educ Open 2024;6:100190. https://doi.org/10.1016/j.caeo.2024.100190.

[62] Tomova L, Andrews JL, Blakemore SJ. The importance of belonging and the avoidance of social risk taking in adolescence. Dev Rev 2021;61:100981. https://doi.org/10.1016/j.dr.2021.100981.

[63] Hubin W, Jinjie X. Efforts to break the 'score determinism' and transfer college enrolment from recruiting 'scores' to 'people': the exploration and practice of comprehensive quality evaluation of general high school students in Shanghai. Educ Philos Theory 2022;54(10):1608–20. https://doi.org/10.1080/00131857.2020.1794152.

[64] Culp, J. (2023). Ethics of education. In Encyclopedia of the philosophy of law and social philosophy (pp. 1–7). https://doi.org/10.1007/978-94-007-6730-0_1055-1.

[65] Bücker S, Nuraydin S, Simonsmeier BA, Schneider M, Luhmann M. Subjective well-being and academic achievement: a meta-analysis. J Res Pers 2018;74:83–94. https://doi.org/10.1016/j.jrp.2018.02.007.

[66] Madsen DØ, Toston DM. ChatGPT and digital transformation: A narrative review of its role in health, education, and the economy. Digital 2025;5(3). https://doi.org/10.3390/digital5030024.

[67] Bossert LN, Loh W. Why the carbon footprint of generative large language models alone will not help us assess their sustainability. Nat Mach Intell 2025:1–2. https://doi.org/10.1038/s42256-025-00979-y.

[68] Alexander, L., & Moore, M. (2024). Deontological ethics. In E. N. Zalta & U. Nodelman (editors), The stanford encyclopedia of philosophy (Winter 2024 Edition).

[69] De Ruyter D, Schinkel A. Ethics education at the university: from teaching an ethics module to education for the good life. Bordon. Rev Pedagog 2017;69(4):125–38. https://doi.org/10.13042/Bordon.2017.690409.

[70] Kasneci E, Seßler K, Küchemann S, Bannert M, Dementieva D, Fischer F, Kasneci G. ChatGPT for good? On opportunities and challenges of large language models for education. Learn Individ Differ 2023;103:102274. https://doi.org/10.1016/j.lindif.2023.102274.

[71] Parry G, Reynoldson C. Creating an authentic learning environment in economics for MBA studies. Cases on the human side of information technology. IGI Global; 2006. p. 76–87. https://doi.org/10.4018/978-1-59140-594-8.ch006.

[72] Dul J. How to sample in necessary condition analysis (NCA). Eur J Int Manag 2024;23(1):1–12. https://doi.org/10.1504/EJIM.2024.138446.

[73] Faul F, Erdfelder E, Buchner A, Lang A-G. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods 2009;41:1149–60. https://doi.org/10.3758/BRM.41.4.1149.

[74] Arias-Oliva M, Pelegrín-Borondo J, Almahameed AA, Andrés-Sánchez JD. Ethical attitudes toward COVID-19 passports: evidences from Spain. Int J Environ Res Public Health 2021;18(24):13098. https://doi.org/10.3390/ijerph182413098.

[75] Reinartz W, Haenlein M, Henseler J. An empirical comparison of the efficacy of covariance-based and variance-based SEM. Int J Res Mark 2009;26(4):332–44. https://doi.org/10.1016/j.ijresmar.2009.08.001.

[76] Dul, J., & Buijs, G. (2025). NCA R package. https://doi.org/10.32614/CRAN.package.NCA.

[77] Rademaker ME, Schuberth F. cSEM: Composite-based structural equation modeling (Version 0.6.0) [R package]. https://CRAN.R-project.org/package=cSEM.

[78] Hair JF, Risher JJ, Sarstedt M, Ringle CM. When to use and how to report the results of PLS-SEM. Eur Bus Rev 2019;31(1):2–24. https://doi.org/10.1108/EBR-11-2018-0203.

[79] Barbosa J. Higher education in ethics and deontology: student's perspective survey. WSEAS Trans Adv Eng Educ 2024;21:128–38.

[80] Gauttier S. I've got you under my skin'–The role of ethical consideration in the (non-) acceptance of insideables in the workplace. Technol Soc 2019;56:93–108. https://doi.org/10.1016/j.techsoc.2018.09.008.

[81] Dahri NA, Yahaya N, Al-Rahmi WM, Vighio MS, Alblehai F, Soomro RB, Shutaleva A. Investigating AI-based academic support acceptance and its impact on students' performance in Malaysian and Pakistani higher education institutions. Educ Inf Technol 2024;29(14):18695–744. https://doi.org/10.1007/s10639-024-12599-x.

[82] Mustofa RH, Kuncoro TG, Atmono D, Hermawan HD, Sukirman. Extending the technology acceptance model: the role of subjective norms, ethics, and trust in AI tool adoption among students. Comput Educ: Artif Intell 2025;8:100379. https://doi.org/10.1016/j.caeai.2025.100379.

[83] Sobaih AEE, Elshaer IA, Hasanein AM. Examining students' acceptance and use of ChatGPT in Saudi Arabian higher education. Eur J Investig Health Psychol Educ 2024;14(3):709–21. https://doi.org/10.3390/ejihpe14030047.

[84] Strzelecki A. To use or not to use ChatGPT in higher education? A study of students' acceptance and use of technology. Interact Learn Environ 2024;32(9):5142–55. https://doi.org/10.1080/10494820.2023.2209881.

[85] Shahzad MF, Xu S, An X, Asif M, Javed I. Do generative AI technologies play a double-edged sword role in education? Findings from hybrid approach using PLS-SEM and fsQCA. Education and Information Technologies; 2025. p. 1–30. https://doi.org/10.1007/s10639-025-13528-2.

[86] Elshaer IA, Hasanein AM, Sobaih AEE. The moderating effects of gender and study discipline in the relationship between university students' Acceptance and use of ChatGPT. Eur J Investig Health Psychol Educ 2024;14(7):1981–95. https://doi.org/10.3390/ejihpe14070132.

[87] Cebrián Cifuentes S, Guerrero Valverde E, Checa Caballero S. The vision of university students from the educational field in the integration of ChatGPT. Digital 2024;4(3):648–59. https://doi.org/10.3390/digital4030032.

[88] Farrelly T, Baker N. Generative artificial intelligence: implications and considerations for higher education practice. Educ Sci 2023;13(11):1109. https://doi.org/10.3390/educsci13111109.

[89] Ma J, Wang P, Li B, Wang T, Pang XS, Wang D. Exploring user adoption of ChatGPT: a technology acceptance model perspective. Int J Hum–Comput Interact 2025;41(2):1431–45. https://doi.org/10.1080/10447318.2024.2314358.

[90] Strzelecki A, ElArabawy S. Investigation of the moderation effect of gender and study level on the acceptance and use of generative AI by higher education students: comparative evidence from Poland and Egypt. Br J Educ Technol 2024;55(3):1209–30. https://doi.org/10.1111/bjet.13425.