Journal of Information, Communication and Ethics in Society

Ethical boundaries in student use of large language models: a necessary condition analysis of two essay-writing uses

Jorge de Andrés-Sánchez
Social and Business Research Lab, Department of Business Administration, Universitat Rovira i Virgili, Tarragona, Spain

Antonio Pérez-Portabella
Department of Communication Studies, Universitat Rovira i Virgili, Tarragona, Spain

Mario Arias-Oliva
Marketing Department, Universidad Complutense de Madrid, Madrid, Spain, and

Manuel Ollé-Sesé
Department of Procedural Law and Criminal Law, Universidad Complutense de Madrid, Madrid, Spain

Received 3 January 2026
Revised 2 March 2026
9 April 2026
Accepted 19 May 2026

Abstract

Purpose – This study aims to examine how university students ethically evaluate two uses of large language models (LLMs) in academic essay writing and, more centrally, whether ethical evaluations operate as empirical bottlenecks for behavioural intention across these scenarios. Using necessary condition analysis (NCA), the paper investigates whether higher intention co-occurs only when ratings on four ethical dimensions exceed minimum levels: moral justice (JU), relativism (RE), consequentialist evaluation (CO) and deontology (DE).

Design/methodology/approach – A sample of 151 students from two Spanish universities evaluated two scenarios: using an LLM to improve an already written essay and using it to generate most of the content. After each scenario, participants reported their behavioural intention and ethical assessments. Paired *t*-tests and repeated-measures multivariate analysis of variance (MANOVA) compared the scenarios, while NCA assessed whether each ethical dimension acted as a necessary condition for intention.

Findings – Ethical evaluations and intention to use LLMs were significantly higher in the editing scenario, indicating that students differentiate between uses with different moral intensities. Ratings in the generative scenario remained moderately positive, and intention slightly exceeded neutrality, revealing ambivalence.

© Jorge de Andrés-Sánchez, Antonio Pérez-Portabella, Mario Arias-Oliva and Manuel Ollé-Sesé. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at http://creativecommons.org/licences/by/4.0/

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).

Conflicts of interest: he authors declare no conflicts of interest.



Journal of Information, Communication and Ethics in Society
Emerald Publishing Limited
1477-996X
DOI 10.1108/JICES-01-2026-0003

Downloaded from www.emerald.com/jices/article-pdf/doi/10.1108/JICES-01-2026-0003/11689032/jices-01-2026-0003en.pdf by guest on 09 November 2026

NCA indicated that, in this sample, low levels of JU, RE and CO were not observed alongside higher intention in either scenario, with less demanding bottlenecks in the generative scenario. DE emerged as a necessary condition only in the editing scenario; in the generative scenario, it did not appear as a bottleneck in the NCA sense.

Research limitations/implications – The convenience sample, self-reported measures and cross-sectional design limit generalisability and causal inference.

Practical implications – Findings underscore the need for differentiated institutional policies, strengthened ethics education and assessment redesign.

Social implications – The framework used here can be adapted to examine LLM use in domains with deeper moral stakes, helping anticipate ethical risks and guide responsible AI adoption in society.

Originality/value – This study extends research on students' ethical evaluations of LLM use by introducing a necessary condition perspective, which is still rare in academic integrity research. Rather than only showing that one scenario is judged more acceptable than the other, the paper identifies which ethical dimensions act as empirical bottlenecks for intention and shows that these bottlenecks vary across scenarios. Within this framework, the asymmetric role of DE emerges as a substantive finding.

Keywords Multidimensional ethics scale, Generative artificial intelligence, Education assessment, Ethics in education, Necessary conditions

Paper type Research paper

1. Introduction

Over the past decade, artificial intelligence (AI) has significantly reshaped the landscape of educational assessment. What began as a tool for automating simple tasks, such as grading multiple-choice exams or organising test logistics, has evolved into something far more dynamic and capable (Arias-Oliva *et al.*, 2026). Large language models (LLMs) have greatly expanded the potential of AI in learning. LLMs facilitate the automated assessment of essays and provide real-time feedback on written work (Bhaskar *et al.*, 2025). These tools have evolved to become interactive systems that engage students through adaptive feedback mechanisms, enabling more individualised and stimulating educational experiences (Kasneci *et al.*, 2023).

While the integration of AI into learning assessment offers significant advantages, it also raises ethical and social concerns. Issues such as algorithmic accountability, transparency, explainability, evaluative fairness, assessment validity, bias, educational equity, institutional accountability and academic integrity must be carefully considered to ensure the ethical implementation of AI in higher education (Arias-Oliva *et al.*, 2026). Addressing these concerns requires the adoption of methodological frameworks that actively involve human oversight to support the responsible development and integration of smart technologies (Bulut *et al.*, 2024).

An important concern is that AI-generated content may facilitate academic misconduct. Consequently, both instructors and students commonly associate the use of LLMs with cheating (Coldwell, 2024). A common assignment format in higher education is the academic essay, which allows educators to assess not only students' understanding of course content but also their ability to integrate information, construct coherent arguments and develop autonomous critical thinking (Aylsworth and Castro, 2024). Its relevance is further justified because essays promote individual reflection, the organisation of complex ideas and the formulation of independent judgements about the subject matter (Dahl *et al.*, 2023). Beyond the erosion of this form of assessment, which aims to cultivate students' autonomy in developing their own thinking with respect to the subject matter, this concern is compounded by the fact that plagiarism detection tools may not be equipped to identify AI-generated text (Arias-Oliva *et al.*, 2026).

Downloaded from www.emerald.com/jices/article-pdf/doi/10.1108/JICES-01-2026-0003/11689032/jices-01-2026-0003en.pdf by guest on 09 November 2026

This study examines whether university students, when deciding to use LLMs as assistants for essay writing, genuinely consider ethical criteria across two scenarios with clearly different moral weight. In the first scenario (Scenario 1), the LLM is used exclusively to refine the presentation of an essay that the student has written independently. In the second scenario (Scenario 2), the student simply submits an essay generated entirely by the LLM without even a minimal review. While the second scenario clearly constitutes academic misconduct, the first is likely to be viewed as more acceptable than full-text generation. Still, its ethical status is not uniform across courses, disciplines, institutions and assessment designs, as it depends on local rules, learning objectives, disclosure norms and whether linguistic refinement is itself part of the competence being assessed. However, although it is also a practice commonly declared by students (Hua and Cunningham, 2025), it is not harmless for a student who is still in training. Academic essay writing is also a form of practice aimed at improving written expression and developing the ability to organise ideas. An overreliance on LLMs may hinder the acquisition of these essential transversal skills (Meyer *et al.*, 2023). Moreover, it is widely recognised that, because of psychological phenomena such as the law of the instrument (Maslow, 1966) and the fact that using LLMs is an online behaviour carried out anonymously – conditions that facilitate cheating (Suler, 2004) – the transition from a use with certain ethical nuances to one that clearly constitutes misconduct becomes significantly easier.

In this regard, we draw on four ethical criteria derived from major currents in moral philosophy to assess the appropriateness of decision-making: moral justice (JU), relativism (RE), consequentialist evaluation (CO) and deontology (DE). These criteria are used to explain students' behavioural intention to use LLMs (BEIN) across two scenarios with different moral implications. They originate from the Multidimensional Ethics Scale (MES), specifically the adaptation by Shawver and Sennetti (2009) of the original scale in Reidenbach and Robin (1990), and have been widely applied to examine ethically controversial academic behaviours such as plagiarism (Prashar *et al.*, 2023; Yang, 2012) and collaborative misconduct during examinations (Jung, 2009; Leonard *et al.*, 2017; Yang, 2012). They have also been used to analyse the influence of ethical judgements in the adoption of emerging and morally controversial technologies, such as cerebral implants (Pelegrín-Borondo *et al.*, 2020a).

Specifically, by using the analytical schema in Figure 1, we pose the following research questions (*RQ*s):

RQ1. Do students perceive the two scenarios as differing in moral intensity?

RQ2. Do minimum ethical evaluations on each criterion function as necessary conditions for intention to use LLMs across the two scenarios?

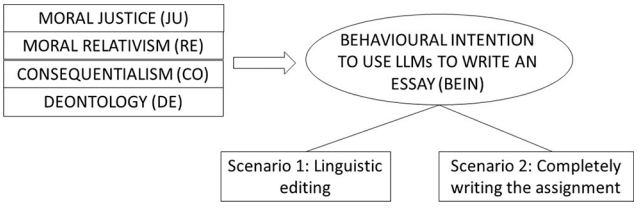


Figure 1. Theoretical groundwork used in this paper

Source: Authors' own work

Downloaded from www.emerald.com/jices/article-pdf/doi/10.1108/JICES-01-2026-0003/11689032/jices-01-2026-0003en.pdf by guest on 09 November 2026

RQ1 establishes whether students meaningfully differentiate between the two uses of LLMs. *RQ2* addresses the paper's central analytical contribution. While *RQ1* is addressed through conventional tests of mean differences, such as analysis of variance and related procedures, *RQ2* implicitly requires a logic of necessity (Dul, 2016). In the MES approach to decision-making, relationships among variables are typically examined using regression-based methods (see, e.g. Prashar *et al.*, 2023) that aim to determine whether "a more positive ethical evaluation on criterion X implies a higher intention to use", which corresponds to a sufficient condition logic. In contrast, we are interested in whether "it is necessary that criterion X be judged acceptable for intention to use to occur", which motivates our use of necessary condition analysis (NCA) (Dul, 2016, 2024a). This design allows us not only to compare the two scenarios, but also to examine whether the ethical bottlenecks associated with intention differ across them.

The structure of the paper is as follows. The next section presents the theoretical perspective, the sample and the analytical methods used in the study. The subsequent section reports the empirical results. Finally, we discuss the key implications and present the most relevant conclusions of the research.

2. Materials and methods

2.1 A multiple ethics scale to evaluate the use of large language models

The outcome evaluated in this study is students' behavioural intention (BEIN) to use LLMs in the two assessed scenarios. Whereas using an LLM to edit a text may raise some ethical concerns, using it to generate the entire assignment is clear misconduct. In this regard, we understand behavioural intention as the predisposition to engage in the behaviours under analysis (Ajzen, 1991). The analytical framework in Figure 1 examines the extent to which a minimum ethical evaluation across four ethical dimensions – JU, moral RE, CO and DE – is a *sine qua non* condition for achieving a predisposition to use LLMs.

JU refers to acting with consideration towards all individuals, particularly in situations involving differences in power, resources or capabilities (Killen, 2018). This principle recognises unequal starting conditions; thus, fairness requires providing each person what they need to access genuine opportunities on equal terms (Arias-Oliva *et al.*, 2026). It deepens the concept of equality – which focuses on uniform treatment – by acknowledging the specific circumstances of each person, with the aim of correcting imbalances and fostering a more just environment (Gosepath, 2021). This approach aligns with care ethics, which holds that GenAI use should be guided by truthfulness, authenticity, respect, equity and social responsibility (Li *et al.*, 2025).

Moral RE holds that actions cannot be classified as universally right or wrong; instead, they must be evaluated in relation to the social, cultural or individual context in which they occur (Reidenbach and Robin, 1990). In education, tools such as ChatGPT are not inherently ethical or unethical; their evaluation depends on context, institutional policies, academic standards and explicit or implicit instructor–student agreements (Licht, 2024). Students' behaviour is undoubtedly influenced by their peers (Strzelecki, 2024) as well as by instructors (Tlais *et al.*, 2025). The importance of the peer cheating effect must also be acknowledged: when students perceive that their peers engage in misconduct, they tend to justify similar behaviour with arguments such as "everyone does it" or "it's normal" (Zhao *et al.*, 2022).

Consequentialism holds that the moral status of an action depends on whether its outcomes generate a positive balance of utility (Sinnott-Armstrong, 2023). Within the MES tradition (Reidenbach and Robin, 1990), this philosophical logic is captured empirically through a CO, which encompasses both utilitarian and egoistic strands of reasoning insofar

Downloaded from www.emerald.com/jices/article-pdf/doi/10.1108/JICES-01-2026-0003/11689032/jices-01-2026-0003en.pdf by guest on 09 November 2026

as moral judgements may be based on perceived benefits or harms to oneself and/or others. Accordingly, CO can be appraised from two complementary perspectives. From an egoistic standpoint, a behaviour is considered morally acceptable when it advances one's own interests, either by securing future benefits or by providing immediate gratification. In contrast, the utilitarian view judges an action as morally appropriate when it contributes positively to collective welfare or to society as a whole (Pérez-Portabella *et al.*, 2025). In the present study, CO is operationalised mainly as perceived outcome favourability, especially in self-relevant academic terms. According to this criterion, a student must weigh whether using an LLM allows them to obtain certain benefits – such as achieving a higher grade or reducing the effort required for an assignment and reallocating that time to more personally valuable activities – against potential drawbacks. These include becoming increasingly dependent on LLMs for tasks such as reports or research notes, introducing erroneous information because of hallucinations (Han *et al.*, 2025) or even being caught in academic fraud if the use of LLMs is explicitly prohibited (Pérez-Portabella *et al.*, 2025).

From a DE perspective, the moral value of an action does not depend on its outcomes but on its conformity with ethical principles, rules and obligations that are established *a priori* (Alexander and Moore, 2024). In the university context, this implies that student behaviour must be evaluated according to its alignment with the moral duties and ethical commitments that govern the relationship between students, educational institutions and society (Pérez-Portabella *et al.*, 2025). The central issue is fulfilling the responsibilities derived from a social contract – explicit or implicit – that determines which practices are morally legitimate in the academic environment (Reidenbach and Robin, 1990). This perspective includes recognising the deontological obligations associated with both one's future profession and civic responsibility (De Ruyter and Schinkel, 2017), actively engaging in one's own learning process – particularly when education is fully or partially funded with public resources – and assuming responsibility for achieving academic performance that authentically reflects this formative effort (Pérez-Portabella *et al.*, 2025).

Accordingly, given that the second scenario presents more ethical problems than the first and in line with *RQ1*:

H1a. Behavioural intention to use LLMs in Scenario 1 is higher than behavioural intention to use LLMs in Scenario 2.

H1b. Ethical evaluations in Scenario 1 are more favourable than those in Scenario 2.

Likewise, regarding the influence of ethical judgements on the intention to use LLMs in the scenarios evaluated, we propose (*RQ2*):

H2a. A positive perception from the perspective of moral justice is necessary for the behavioural intention to use LLMs in a given scenario.

H2b. A positive perception from the perspective of moral relativism is necessary for the behavioural intention to use LLMs in a given scenario.

H2c. A positive consequentialist evaluation, operationalised here mainly as perceived outcome favourability, is necessary for the behavioural intention to use LLMs in a given scenario.

H2d. A positive perception from the perspective of deontology is necessary for the behavioural intention to use LLMs in a given scenario.

Downloaded from www.emerald.com/jices/article-pdf/doi/10.1108/JICES-01-2026-0003/11689032/jices-01-2026-0003en.pdf by guest on 09 November 2026

2.2 Variable measurement

The survey consisted of a self-administered, anonymous electronic questionnaire, a format expected to reduce social desirability bias. The instrument started by presenting the two scenarios (Scenario 1 and Scenario 2).

After each scenario, students were asked to indicate their predisposition to use the LLM in that situation and to provide an ethical evaluation of that use with the items reported in Items used in this paper.

With regard to the proposed vignette, evaluate the following statements from 0 to 10, where 0 implies complete disagreement and 10 complete agreement.

Behavioural intention (BEIN):

- BEIN1: I intend to use LLMs in this scenario.
- BEIN2: I predict that I will use LLMs in this scenario.

Moral justice (JU):

- JU1 = The use of LLMs in this scenario is fair.
- JU2 = The use of LLMs in this scenario is equal.
- JU3 = The use of LLMs in this scenario is right.

Relativism (RE):

- RE1 = The use of LLMs in this scenario is acceptable to my peers.
- RE2 = The use of LLMs in this scenario is acceptable in my environment.
- RE3 = The use of LLMs in this scenario is acceptable to people whose opinion I respect.

Consequentialist evaluation (CO):

- CO1 = The use of LLMs in this scenario will provide relevance and prestige.
- CO2 = The use of LLMs in this scenario is rewarding.
- CO3 = The use of LLMs in this scenario is useful.
- CO4 = The use of LLMs in this scenario has good cost–benefit balance.

Deontology (DE):

- DE1 = The use of LLMs in this scenario respects an implicit contract with my environment/society.
- DE2 = The use of LLMs in this scenario aligns with what is expected of me as a student.

BEIN was measured using the scale by Venkatesh and Davis (2000), used in MES behavioural research (Pelegrín-Borondo *et al.*, 2020). For the assessment of ethical judgements, the study used the MES in Shawver and Sennetti (2009), in which the subscales of egoism and utilitarianism are combined into a single dimension of CO (De Andrés-Sánchez, 2025). CO was measured using four items drawn from prior MES operationalisation of egoism and utilitarianism (Shawver and Sennetti, 2009). Consistent with the MES CO dimension, these items capture perceived outcome favourability as a basis for moral evaluation, encompassing both utilitarian and egoistic strands. Specifically, CO reflects self-relevant academic benefits (e.g. “relevance and prestige”, “rewarding”) as well as instrumental assessments of outcomes (e.g. “useful”, “good cost–benefit balance”).

Downloaded from www.emerald.com/jices/article-pdf/doi/10.1108/JICES-01-2026-0003/11689032/jices-01-2026-0003en.pdf by guest on 09 November 2026

Accordingly, in this study, CO is best understood as perceived outcome favourability, primarily in self-regarding academic terms, rather than as a full normative consequentialist appraisal that weighs harms and benefits for all stakeholders.

Items used in this paper shows that BEIN and DE were measured with two items. While short scales can increase random measurement error and limit content coverage, two-item measures are common in vignette-based designs and are consistent with prior MES-based operationalisation. Moreover, the selected items were chosen to capture the most central facets of each construct within the study's scenario framing, prioritising respondent burden and comparability across scenarios (Pelegrín-Borondo *et al.*, 2020; Prashar *et al.*, 2023; Shawver and Sennetti, 2009).

All questions were answered on a 0–10 scale, rather than the more common but narrower 5- or 7-point scales. This choice was made for two reasons. First, it is a highly intuitive response format in many countries, where evaluations on this range are common, and 5 represents the neutral point. Second, this type of scale captures nuances that respondents can discriminate more accurately than with shorter scales and tends to approximate a normal distribution more closely (Bisquerra Alzina and Pérez-Escoda, 2015). For these reasons, the 0–10 scale has been previously used in the evaluation of ethical judgements through the MES, for example, in studies on the use of brain implants (Pelegrín-Borondo *et al.*, 2020a) or in the assessment of new sports technologies (De Andrés-Sánchez, 2025).

2.3 Data collection and participants

Data collection took place between April and June 2025 among university students enrolled in three specific degree programs often grouped under the broad label of "Social Sciences", Social Work, Business Administration and Management and Economics, at two Spanish universities. These programs combine essay-oriented coursework typically associated with the humanities (e.g. philosophy/history and related ethical–normative content) with quantitative training (e.g. statistics/economics), and essay-based assignments are common (Correa *et al.*, 2024). Likewise, within these degree contexts, the academic use of GenAI by students is already widespread, employed both for linguistic support and for information retrieval (Farinosi and Melchior, 2025). Accordingly, this sampling frame was considered suitable for examining ethical evaluations of LLM use in essay-based assessment contexts.

Participants were selected through purposive sampling. While this non-probability sampling strategy limits statistical generalisability, the study is intended as an analytical examination of ethical evaluation patterns within a clearly defined academic context. The voluntary, anonymous and self-administered survey was distributed via hyperlink. A total of 151 valid responses were obtained, with the sample distributed as follows:

- Gender: 43 males and 108 females.
- Age: Between 18 and 48 years; mean age = 21.89 years, SD = 3.64.
- Academic performance: 44 students above the mean and 107 at or below the mean.
- Employment status: 58 full-time workers, 38 part-time workers and 55 not employed.

2.4 Data analysis

Before testing the hypotheses proposed in Section 2.1, we assessed the internal validity of the scales using the standard indicators: Cronbach's alpha, average variance extracted (AVE) and factor loadings. Measurement invariance across scenarios was assessed via confirmatory factor analysis (CFA) by treating Scenario 1 and Scenario 2 as two groups for invariance-testing purposes (configural, metric and scalar models). Following commonly used

Downloaded from www.emerald.com/jices/article-pdf/doi/10.1108/JICES-01-2026-0003/11689032/jices-01-2026-0003en.pdf by guest on 09 November 2026

guidelines, invariance was evaluated primarily through changes in approximate fit indices rather than χ^2 differences, using thresholds of change in comparative fit index (ΔCFI) ≤ 0.010 and change in root mean square error of approximation (ΔRMSEA) ≤ 0.015 (Chen, 2007) and considering change in standardized root mean square residual (ΔSRMR) as supplementary information.

The evaluation of *H1a* and *H1b* was conducted using two complementary approaches:

(1) A paired-samples Student's *t*-test was applied for each item.

(2) For each construct, a within-subjects repeated-measures multivariate analysis of variance (MANOVA) was used.

Regarding *H2a–H2d*, their testing was carried out using NCA. For both the dependent variable (BEIN) and the explanatory variables (JU, RE, CO and DE), the arithmetic mean of the corresponding items was calculated. To estimate the free disposal hull, the ceiling envelopment technique was applied.

Several R packages were used to conduct these analyses. Internal reliability and validity of the scales (Cronbach's alpha, AVE and factor loadings) were assessed using the psych and lavaan packages. Paired *t*-tests by item were performed using base R functions and the stats package, while the repeated-measures MANOVA was implemented with car and stats. Finally, to test the hypotheses related to NCA (*H2a–H2d*), we used the NCA package.

The adequacy of the sample size was evaluated considering the statistical procedures used. According to G*Power 3.1 (Faul *et al.*, 2009), for the paired-samples *t*-test, the sample size enables the detection of small effects (Cohen's $d = 0.25$) with 80% power and a 5% significance level. In the case of the within-subjects MANOVA, the available sample allows detecting effect sizes between $f = 0.12$ (constructs with two items) and $f = 0.10$ (constructs with four items), also classified as small, with 80% power and $\alpha = 0.05$.

In NCA, the capacity of the sample size to yield meaningful results can also be evaluated (Dul, 2024b). Using the R NCA package, we verified that the sample size was adequate to detect effects with empirical relevance ($d > 0.10$; Dul, 2016). Assuming a uniform marginal distribution of the joint distribution function BEIN–ethical construct, the estimated power is 100% under a 5% significance level.

3. Results

3.1 Results for the RQ1

Table 1 shows that all scales exhibit adequate internal consistency, with Cronbach's alpha and composite reliability ranging from 0.70 to 0.95. They also show convergent reliability,

Table 1. Measures of internal scale validity

	Scenario 1			Scenario 2		
Variable	CA	CR	AVE	CA	CR	AVE
BEIN	0.900	0.913	0.839	0.938	0.939	0.884
JU	0.856	0.857	0.667	0.844	0.860	0.682
RE	0.828	0.851	0.659	0.880	0.894	0.743
CO	0.873	0.884	0.656	0.908	0.912	0.721
DE	0.865	0.868	0.767	0.943	0.944	0.893

Note(s): CA = Cronbach's alpha, CR = composite reliability, AVE = average variance extracted
Source(s): Authors' own work

Downloaded from www.emerald.com/jices/article-pdf/doi/10.1108/JICES-01-2026-0003/11689032/jices-01-2026-0003en.pdf by guest on 09 November 2026

Downloaded from www.emerald.com/jices/article-pdf/doi/10.1108/JICES-01-2026-0003/11689032/jices-01-2026-0003en.pdf by guest on 09 November 2026

Table 2. Descriptive statistics of items and pairwise tests of mean differences between evaluations in Scenarios 1 and 2

	Scenario 1			Scenario 2			Mean differences in Scenarios 1 and 2			
Variable	Mean	SD	FL	Mean	SD	FL	Diff	*t*-Ratio	*p*-value	Cohen's d
Behavioural intention BEIN										
BEIN1	6.94	2.94	0.959	5.44	2.92	0.971	1.50	6.13	<0.001	0.499
BEIN2	7.51	2.35	0.959	5.99	2.75	0.971	1.52	7.55	<0.001	0.614
Moral justice (JU)										
JU1	7.6	1.98	0.891	4.84	2.76	0.927	2.76	12.14	<0.001	0.988
JU2	7.71	2.16	0.876	6.74	2.69	0.760	0.97	5.34	<0.001	0.435
JU3	7.04	2.01	0.879	4.81	2.92	0.926	2.23	9.71	<0.001	0.79
Relativism (RE)										
RE1	7.83	1.97	0.890	6.12	2.86	0.889	1.71	8.98	<0.001	0.731
RE2	7.62	1.95	0.911	5.63	2.89	0.958	1.99	9.11	<0.001	0.741
RE3	6.33	2.58	0.818	4.3	3.08	0.852	2.03	8.93	<0.001	0.726
Consequentialist evaluation (CO)										
CO1	8.05	1.97	0.888	6.81	2.71	0.881	1.24	6.74	<0.001	0.548
CO2	6.6	2.62	0.817	4.96	3.18	0.842	1.64	6.74	<0.001	0.549
CO3	8.4	1.87	0.875	6.81	2.8	0.921	1.59	8.15	<0.001	0.663
CO4	7.29	2.07	0.862	5.99	2.88	0.907	1.3	5.92	<0.001	0.482
Deontology (DE)										
DE1	6.17	2.45	0.940	4.52	3.1	0.973	1.65	7.38	<0.001	0.600
DE2	5.81	2.73	0.940	4.31	3.16	0.973	1.5	6.86	<0.001	0.558

Note(s): SD = standard deviation, FL = factor loading, Diff = difference between the item means in Scenario 1 and Scenario 2
Source(s): Authors' own work

because AVE values exceed 0.50 and all item factor loadings in Table 2 are above 0.70. In terms of measurement invariance, we tested whether the measurement model was equivalent across the two scenarios by estimating a sequence of multi-group CFA models (configural, metric and scalar). Overall, the results showed very small changes in fit across increasingly constrained models, with ΔCFI = −0.003 from the configural to the metric model and ΔCFI = −0.005 from the metric to the scalar model; RMSEA remained essentially stable (ΔRMSEA = −0.003 and −0.001, respectively), and SRMR increased slightly (ΔSRMR = 0.014 and 0.005, respectively). SRMR increased slightly at the metric step; however, ΔCFI and ΔRMSEA remained well within recommended thresholds, supporting approximate invariance. Thus, these results indicate that the constructs were measured in a largely comparable way across scenarios, supporting meaningful comparisons between scenarios. Importantly, these conclusions were not materially affected by the inclusion of two-item constructs (BEIN and DE), which showed adequate reliability and behaved consistently with the overall measurement model.

Table 2 shows that BEIN values in the first scenario are around or above the scale midpoint, whereas in the second scenario, they are notably lower – although still above the neutral value of 5. These results indicate greater acceptance of LLM use in the first scenario and more moderate approval in the second. Table 2 also shows that the mean difference between BEIN items across both scenarios is significant ($p < 0.001$), with Cohen's d effect sizes ranging from small but near-medium effects (BEIN1, $d = 0.499$) to medium effects (BEIN2, $d = 0.614$). MANOVA analysis confirms that the differences in BEIN between the two scenarios are significant (Pillai's trace = 0.266, $p < 0.001$), with an effect size of Cohen's $f = 0.603$, considered large (Table 3).

Table 2 shows that ethical evaluations of LLM use in the first scenario are systematically higher than in the second. In the first scenario, JU, RE and CO scores typically exceed 7 (except RE3 and CO2), reaching even close to 8 (CO1 and CO3). In the second scenario, scores range from 4.30 (RE3) to 6.81 (CO1 and CO3). In both scenarios, the lowest evaluations correspond to DE: around 6 in the first scenario and not exceeding 5 in the second. The mean differences test indicates that all ethical evaluations of LLM use are significantly more favourable in the first scenario. Cohen's d effect sizes range from small-to-medium effects (JU2, $d = 0.435$) to large effects (JU1, $d = 0.988$). The overall pattern suggests medium effects, for example, $d = 0.548$ (CO1) and $d = 0.741$ (RE2).

Regarding the MANOVA results shown in Table 3, differences between both scenarios are significant for all ethical constructs and exhibit large Cohen's f effect sizes. For JU, Pillai's trace = 0.439, $p < 0.001$, $f = 0.885$; for RE, Pillai's trace = 0.413, $p < 0.001$, $f = 0.838$; for CO, Pillai's trace = 0.301, $p < 0.001$, $f = 0.657$; and for DE, Pillai's trace = 0.276, $p < 0.001$, $f = 0.617$.

Table 3. Results of the MANOVA test of differences in the constructs between the two scenarios

Variable	Pillai trace	Wilks	Hotelling	Roy	Snedecor's F	*p*-value	Cohen's *f*
BEIN	0.266	0.734	0.363	0.363	54.49	<0.001	0.603
JU	0.439	0.561	0.784	0.784	117.54	<0.001	0.885
RE	0.413	0.587	0.704	0.704	105.62	<0.001	0.839
CO	0.301	0.699	0.431	0.431	64.66	<0.001	0.657
DE	0.276	0.724	0.381	0.381	57.11	<0.001	0.617

Source(s): Authors' own work

Downloaded from www.emerald.com/jices/article-pdf/doi/10.1108/JICES-01-2026-0003/11689032/jices-01-2026-0003en.pdf by guest on 09 November 2026

Table 4. Effect sizes as necessary conditions of ethical constructs

Variable	Scenario 1 Effect size (d)	Scenario 1 *p*-value	Scenario 2 Effect size (d)	Scenario 2 *p*-value
JU	0.340	<0.001	0.200	<0.001
RE	0.169	<0.001	0.188	<0.001
CO	0.368	<0.001	0.229	<0.001
DE	0.235	<0.001	<0.001	>0.999

Source(s): Authors' own work

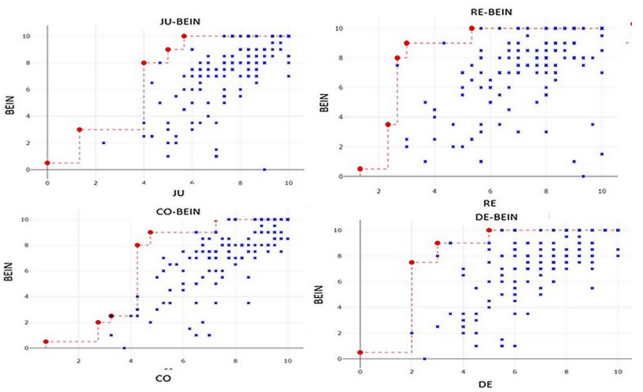


Figure 2. Ceiling envelopment-free disposal hull in Scenario 1
Source: Authors' own work

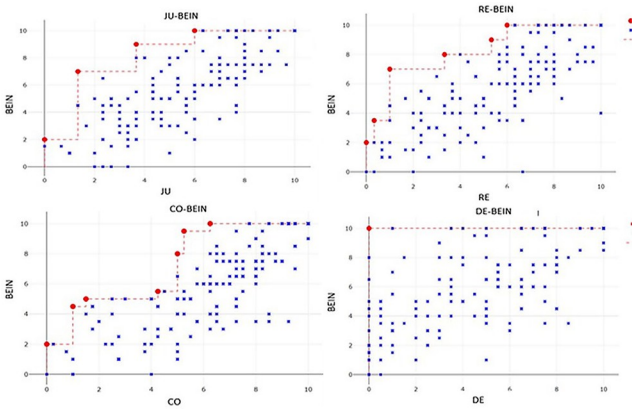


Figure 3. Ceiling envelopment-free disposal hull in Scenario 2
Source: Authors' own work

Downloaded from www.emerald.com/jices/article-pdf/doi/10.1108/JICES-01-2026-0003/11689032/jices-01-2026-0003en.pdf by guest on 09 November 2026

3.2 Results for the RQ2

Table 4 presents the NCA effect sizes for the two scenarios, derived from the ceiling envelopment-free disposal hull shown in Figure 2 (Scenario 1) and Figure 3 (Scenario 2). In Scenario 1, a minimally positive ethical evaluation is necessary for the formation of BEIN. Effect sizes range from large ($d > 0.30$) for JU and CO, to medium ($d > 0.10$) for RE and DE, following the typology proposed by Dul (2016). All effect sizes are statistically significant ($p < 0.001$).

In Scenario 2, the effect sizes of the ethical constructs are noticeably smaller: JU and CO decrease from large to medium effects ($0.10 \leq d \leq 0.30$), although they remain statistically significant ($p < 0.001$), and RE shows a slight increase, also significant ($p < 0.001$). In contrast, DE – which showed a medium effect in Scenario 1 ($d = 0.235$) – is no longer significant in Scenario 2 ($d < 0.001$; $p > 0.999$), indicating that DE does not act as a necessary condition in this case.

Table 5 presents the bottleneck tables for both scenarios, calculated using the mean scores of the items in each construct (JU, RE, CO and DE). In Scenario 1, even very low levels of BEIN (1 / 10) require a minimum ethical level above 1 across all constructs. A neutral BEIN score of 5 requires values close to 4 in JU and CO and above 2 in RE and DE. This pattern persists for intermediate and high BEIN levels (8–9), which require minimum values of approximately 4–5 in JU and CO, while the thresholds for RE and DE remain below 3. Maximum acceptance (BEIN = 10) requires surpassing level 5 in JU, RE and CO – with CO presenting an especially high threshold (7.25) – and reaching level 5 in DE.

In contrast, Table 5 shows that the necessary thresholds in Scenario 2 are far less demanding. First, DE does not act as a necessary condition in any case. Moreover, low BEIN levels (0–2) can be reached without any ethical requirement. Starting from BEIN = 3, modest thresholds emerge for JU, RE and CO, which remain very low (slightly above 1 out of 10) up to BEIN = 5. Requirements for JU and RE remain below 4 even for high BEIN levels (up to 9 in the case of JU and 8 in RE). CO begins to show more demanding thresholds from BEIN = 6 onward, where a minimum level of 5 is required. Finally, achieving the maximum BEIN level (10) requires JU, RE and CO to exceed 6. This pattern reflects the characteristic logic of NCA: higher levels of intention require increasing – and non-substitutable – levels of the necessary conditions. Overall, the results indicate that in Scenario 2 the ethical constraints necessary for forming BEIN are more lenient and gradual than in Scenario 1, while DE does not constitute a bottleneck for intention, in spite of its possible relevance in other analytical approaches.

In conclusion, while *H2*a, *H2*b and *H2*c are fully supported – meaning that a positive evaluation of JU, RE and CO constitutes a necessary condition for the use of LLMs in both scenarios – *H2d*, concerning whether a positive evaluation of DE is necessary for the intention to use LLMs, is confirmed only in the scenario involving the use of an LLM as a text editor, but not in the scenario involving the generation of a full essay.

4. Discussion

4.1 General considerations

This study examines whether university students distinguish the ethical implications of two uses of LLMs in academic essay writing and whether their intention to use these tools is associated with minimal ethical evaluations in each scenario. In Scenario 1, the LLM is used to edit and linguistically refine an essay already written by the student; in Scenario 2, the LLM is used to generate the entire essay.

The analysis is based on the premise that no single criterion determines whether a decision is ethical. To capture this plurality, the study adopts the MES (Shawver and Sennetti, 2009), evaluating four ethical dimensions: JU, RE, CO and DE.

Downloaded from www.emerald.com/jices/article-pdf/doi/10.1108/JICES-01-2026-0003/11689032/jices-01-2026-0003en.pdf by guest on 09 November 2026

Downloaded from www.emerald.com/jices/article-pdf/doi/10.1108/JICES-01-2026-0003/11689032/jices-01-2026-0003en.pdf by guest on 09 November 2026

Table 5. Bottleneck tables for ethical constructs on intention to use in the analysed scenarios

Scenario 1					Scenario 2				
BEIN (Y)	JU	RE	CO	DE	BEIN (Y)	JU	RE	CO	DE
0	NN	NN	NN	NN	0	NN	NN	NN	NN
1	1.333	2.333	2.75	2	1	NN	NN	NN	NN
2	1.333	2.333	2.75	2	2	NN	NN	NN	NN
3	1.333	2.333	4.25	2	3	1.333	0.333	1	NN
4	4	2.667	4.25	2	4	1.333	1	1	NN
5	4	2.667	4.25	2	5	1.333	1	1.5	NN
6	4	2.667	4.25	2	6	1.333	1	5	NN
7	4	2.667	4.25	2	7	1.333	1	5	NN
8	4	2.667	4.25	3	8	3.667	3.333	5	NN
9	5	3	4.75	3	9	3.667	5.333	5.25	NN
10	5.667	5.333	7.25	5	10	6	6	6.25	NN

Note(s): (a) Bottleneck tables are calculated on actual values, which range from 0 to 10. (b) NN stands for non-necessary condition
Source(s): Authors' own work

The results show that students demonstrate clear ethical sensitivity in their use of LLMs. Both their behavioural intention (BEIN) and their ethical evaluations are significantly more positive in Scenario 1 than in Scenario 2, confirming that students perceive the ethical appropriateness of LLM use as context-dependent, as emphasised in the literature on educational applications of LLMs (Kasneci *et al.*, 2023), including their use in the production of academic essays (Gorrieri, 2025). Ethical evaluations in Scenario 1 are relatively high, suggesting that students perceive only weakly some of the ethical concerns also present in this context, such as overreliance on LLMs during training and inequalities in access to AI tools. Strikingly, ethical evaluations in Scenario 2 are also at moderate levels (often above 4 and even above 5 in several items), and behavioural intention – even when it involves generating a complete essay – slightly exceeds the neutral point. This suggests that, in spite of recognizing the ethically problematic nature of full generative use, some students would consider using LLMs if they perceive it as necessary, and may justify such use as a way of "levelling the playing field" depending on their personal circumstances (Pérez-Portabella *et al.*, 2025).

The most distinctive contribution of this study emerges from the application of NCA. Traditionally, analyses based on the MES rely on regression techniques to evaluate whether more positive ethical evaluations translate into stronger behavioural intentions. This logic of sufficiency has dominated seminal contributions (Reidenbach and Robin, 1990), studies in business ethics (Shawver and Sennetti, 2009), analyses of controversial technologies (Pelegrín-Borondo *et al.*, 2020) and research on morally contentious academic behaviours (Prashar *et al.*, 2023). Unlike regression, NCA identifies empirical bottlenecks in the observed data; that is, regions where higher intention does not co-occur with low ethical evaluations. These "necessary" patterns are sample- and operationalisation-bound and do not, by themselves, establish causal necessity or psychological mechanisms. The results show that, in Scenario 1, all four ethical criteria operate as necessary conditions. JU and CO emerge as strong necessary conditions, indicating that, within this sample and vignette framing, high intention is not observed when JU and CO evaluations are very low – i.e. when respondents perceive limited fairness JU and limited CO.

Importantly, NCA detects empty space in the X–Y scatter and expresses it as empirical necessity. Such empty space can reflect multiple mechanisms beyond causal necessity, including ceiling/floor effects, restricted range, common-method variance, vignette-induced demand or omitted variables that jointly shape ethical evaluations and intention. Accordingly, we interpret the identified bottlenecks as descriptive constraints in this data set rather than as evidence of psychological requirements.

In Scenario 1, acceptance of LLM use is highest when respondents view the practice as reasonably fair, for example, when they perceive it as broadly accessible and as helping to level opportunities in assignments where the aim is to demonstrate subject mastery rather than language skills. The NCA pattern for RE indicates that higher intention does not co-occur with very low perceived contextual acceptability, at least as judged by peers or instructors (Strzelecki, 2024; Tlais *et al.*, 2025). Likewise, the pattern for CO suggests that higher intention is unlikely when perceived outcome favourability is very low, that is, when respondents perceive limited usefulness and an unfavourable cost–benefit balance and/or low self-relevant benefits (Chen, 2025), and when LLM use is not seen as conducive to well-being (Hasanein and Sobaih, 2023). The relevance of DE is consistent with the idea that, in Scenario 1, intention to use LLMs co-occurs with the desire to present the best possible assignment. Family expectations and institutional standards may create implicit obligations that motivate students towards better academic performance (Chen, 2023).

Downloaded from www.emerald.com/jices/article-pdf/doi/10.1108/JICES-01-2026-0003/11689032/jices-01-2026-0003en.pdf by guest on 09 November 2026

Scenario 2 presents a markedly weakened ethical structure. JU, RE and CO remain necessary conditions, but with moderate effect sizes and much lower thresholds. Most importantly, DE is not a necessary condition in this scenario: low DE scores did not form an empirical bottleneck for higher intention in this data set. The fact that JU, RE and CO remain necessary conditions, and that many students exceed thresholds to induce a non-trivial BEIN, may be consistent with the possibility that some students might justify misuse of LLMs under certain circumstances. From the perspective of JU, students may view such behaviours not as attempts to cheat but as responses to structural injustices or extreme pressure (Rosen and Dreskin, 2024). From the relativistic perspective, even when instructors disapprove of improper LLM use, peer behaviour decisively influences the likelihood of academic misconduct (Perry *et al.*, 2025); when misuse becomes normalised, dishonesty is justified through social observation and group pressure (Zhao *et al.*, 2022). In a CO setting, the difficulty of distinguishing between AI-generated text and merely edited text (Arias-Oliva *et al.*, 2026) reduces perceived risk, and a low probability of detection is a well-known driver of academic misconduct (Miles *et al.*, 2022).

The asymmetric role of DE is a notable insight. In Scenario 1, where the practice may be seen as acceptable, duty-based considerations seem to act as a minimal moral "gatekeeper" for higher intention, consistent with an implicit academic "social contract". In Scenario 2, however, DE did not emerge as a necessary condition in the NCA sense, as low DE ratings did not form an empirical bottleneck for higher intention. Rather than concluding that DE becomes psychologically irrelevant, we interpret this as a measurement-level pattern: under this scenario framing, duty-based judgements were not the limiting factor for higher intention, consistent with behavioural ethics work emphasising situational pressures and perceived consequences once wrongdoing is contemplated.

4.2 Theoretical implications

This study makes several theoretical contributions to research on ethical decision-making in academic contexts. First, it shows that MES provides a robust framework for analysing students' evaluations of AI-mediated academic practices. More centrally, by applying NCA, the study introduces a logic of necessity that is rarely used in academic integrity research. Rather than asking whether more ethical evaluations correlate with BEIN or simply showing that the two scenarios are judged differently, it identifies the minimum ethical levels required for intention to emerge, shifting attention from sufficiency to moral constraints. In this sense, the paper's main theoretical contribution lies in showing that ethical evaluations can be analysed as scenario-specific empirical bottlenecks for behavioural intention.

The findings also demonstrate that ethical reasoning is scenario-dependent: responses differed systematically across scenarios, consistent with scenario-dependent ethical evaluation; in NCA terms, the estimated bottlenecks were more demanding in Scenario 1 than in Scenario 2. Thus, the study shows not only that the scenarios are evaluated differently, but also that the ethical bottlenecks associated with intention vary across scenarios. This supports situational and contextual theories of moral judgement and extends them to LLM-mediated practices. Finally, the results suggest an asymmetric role of DE: duty-based evaluation appeared as an NCA bottleneck in the editing scenario but not in the generative scenario. This asymmetry may reflect genuine shifts in justification across contexts, but it may also be amplified by vignette clarity and distributional constraints; we therefore avoid interpreting it as evidence that deontological reasoning "switches off" in general.

Downloaded from www.emerald.com/jices/article-pdf/doi/10.1108/JICES-01-2026-0003/11689032/jices-01-2026-0003en.pdf by guest on 09 November 2026

4.3 Practical implications

The findings in this study offer practical implications for universities integrating LLMs into their assessment systems that should be interpreted in light of the purposive nature of the sample (social science students in Spain), which may limit generalisability. First, students do not perceive all forms of AI assistance in the same way: they are more permissive when the tool enhances their own writing but adopt a more ethically ambivalent stance when it replaces their authorship. Institutions can use this insight to develop nuanced policies that differentiate between editing, feedback, idea generation support and full text generation. At the same time, what counts as acceptable editing support is not fixed across higher education contexts but depends on course objectives, disciplinary expectations, institutional rules and disclosure requirements. Policies that treat all uses uniformly risk losing credibility among students (Arias-Oliva *et al.*, 2026). When LLM use is perceived as partially legitimate (e.g. editing support), duty-based guidance centred on responsibilities, disclosure and academic expectations may reinforce responsible use. By contrast, in clearly generative contexts, fairness perceptions, social norms and CO may be more salient in constraining intention. Accordingly, interventions may work best as a combined strategy: clear rules and disclosure requirements, complemented by explicit communication about fairness impacts, peer norms and practical consequences.

Because JU, RE and CO emerge as necessary conditions even in the most ethically problematic scenario, these dimensions may be especially relevant when designing educational interventions for contexts similar to those studied here. Highlighting the inequities created by the improper use of LLMs – for example, how unfair advantages undermine peers' efforts – may deter misconduct more effectively than reminders focused solely on rule compliance. Evidence shows that ethical awareness interventions can heighten students' sensitivity to fairness-based judgements (Prashar *et al.*, 2023). Because RE also shapes BEIN, institutions may benefit from cultivating visible norms of integrity and shared expectations among students. Likewise, because CO functions as a necessary condition even when students contemplate clearly inappropriate uses of LLMs, discussing the real risks of indiscriminate AI use, such as cognitive dependence or hallucinations, may be relevant when addressing students' decisions in contexts similar to those studied here. Advancing the capacity to detect AI-generated content is equally crucial, as perceived impunity fuels academic misconduct with LLMs.

DE emerges as a necessary condition only in Scenario 1. Policies grounded in obligations, duties and expectations may, therefore, be especially useful when behaviour lies in ambiguous or semi-permitted zones. Students already considering misconduct are unlikely to be swayed by rule-based messages alone, so institutions may need to complement normative communication with context- and consequence-based educational strategies. Institutions may consider defining clear criteria for permitted GenAI use in essays, including disclosure of how and why it was used, akin to research writing standards (Sebastião *et al.*, 2025). Finally, the study highlights the need to redesign assessment strategies in light of LLM capabilities. Traditional essays, increasingly vulnerable to misuse, may need to be complemented with oral expositions, process-oriented evaluations, draft portfolios, in-class writing, individualised prompts or AI-aware rubrics. Recognising the legitimate role of LLMs – particularly for editing and linguistic support – may also promote more responsible student behaviour.

5. Conclusions

5.1 Main takeaways

This study demonstrates that students clearly distinguish between ethically acceptable and unacceptable uses of LLMs in academic essay writing. Scenario 1 – editing an already written text – is judged more positively across all ethical dimensions, and students express

Downloaded from www.emerald.com/jices/article-pdf/doi/10.1108/JICES-01-2026-0003/11689032/jices-01-2026-0003en.pdf by guest on 09 November 2026

higher BEIN. Scenario 2 – having the LLM generate much of the essay – receives substantially lower ethical evaluations, although some students still express willingness to engage in this form of misconduct.

The main contribution of this paper is to show that ethical evaluations can function as empirical bottlenecks for students' intention to use LLMs for academic purposes, and that these bottlenecks vary across scenarios. The estimated NCA bottlenecks differed by scenario, indicating that the empirical ethical constraints associated with intention are context-dependent within the present vignette design. Across both scenarios, JU, contextual acceptability (relativism) and outcome favourability within the consequentialist dimension acted as empirical minimum conditions in the NCA sense. By contrast, DE emerged as a bottleneck only in the editing scenario; in the generative scenario, it did not appear as an NCA constraint, a pattern that may reflect vignette salience, ceiling/floor effects or other design- and measurement-related factors. Accordingly, the findings support considering differentiated policies that distinguish between editing-oriented and generative uses of LLMs, combining clear guidance with interventions that emphasise fairness, social norms and practical consequences rather than relying exclusively on rule-based prohibitions (Arias-Oliva *et al.*, 2026).

5.2 Limitations and future research

The study acknowledges several limitations. This study relied on purposive, non-probability sampling from students enrolled in three degree programmes (Social Work, Business Administration and Management and Economics) at two Spanish universities. Accordingly, the findings should be interpreted as context-specific and may not generalise to STEM fields, other institutional types or international settings. Future research should replicate the design across disciplines and countries (and, where feasible, use probability-based recruitment) to test the robustness and boundary conditions of the observed patterns. In addition, the study assesses self-reported behavioural intention rather than observed behaviour; experimental designs, behavioural tasks or digital trace/tracking approaches could examine whether the same necessary condition relationships hold in real-life LLM use.

While consistent with prior MES applications, our CO items focus mainly on outcome favourability in self-relevant academic terms rather than a comprehensive other-regarding consequentialist calculus. Future studies could broaden this by capturing other-regarding consequences (e.g. perceived harm to peers, cohort-level fairness, or impacts on academic integrity) to reflect utilitarian considerations more directly. In addition, we did not directly measure perceived moral intensity; instead, we rely on the vignette design and the observed differences across scenarios as indirect support for the intended manipulation. Future research can strengthen the scenario logic by including a brief measure of perceived moral intensity (e.g. perceived seriousness, harm and social consensus).

Moreover, only two scenarios are analysed, editing support and full text generation, though students often use LLMs in intermediate ways (e.g. brainstorming, paraphrasing, or rewriting). In addition, the ethical meaning of "editing support" may vary across disciplines, courses and institutions depending on local rules, assessment objectives and disclosure expectations. Future studies should explore a broader range of uses to capture how ethical reasoning varies along a continuum. Fourth, the MES framework focuses on four ethical dimensions, but other factors such as moral disengagement, academic self-efficacy and AI literacy may also shape intentions. Integrating these with NCA or combining methods (e.g. structural equation modelling or fuzzy set qualitative comparative analysis) could yield a more comprehensive understanding of behaviour.

Downloaded from www.emerald.com/jices/article-pdf/doi/10.1108/JICES-01-2026-0003/11689032/jices-01-2026-0003en.pdf by guest on 09 November 2026

Notice that NCA identifies ethical thresholds at a single point in time. Longitudinal research is needed to examine how these thresholds evolve as students gain experience with LLMs and as institutional norms change. Given the rapid spread of generative AI in education, understanding how these ethical conditions evolve will be crucial for designing adaptive and forward-looking policies.

Data availability

The data supporting the analysis is available upon reasonable request.

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_20250710_10_SOC).

Informed Consent Statement

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

References

Ajzen, I. (1991), "The theory of planned behavior", *Organizational Behavior and Human Decision Processes*, Vol. 50 No. 2, pp. 179-211 doi: 10.1016/0749-5978(91)90020-T.

Alexander, L., and Moore, M. (2024), "Deontological ethics", In Zalta, E.N. and Nodelman, U. (Eds), *The Stanford Encyclopedia of Philosophy*, (Winter 2024 Edition) Metaphysics Research Lab, Stanford University, available at: https://plato.stanford.edu/archives/win2024/entries/ethics-deontological/.com

Arias-Oliva, M., Pérez-Portabella, A., Ollé Sesé, M., and de Andrés Sánchez, J. (2026), "Rethinking educational assessment in the age of artificial intelligence", In M. and D.A.-H. and S.N. Alvarez, Isabel and Arias-Oliva (Eds), *Ethical and Social Impacts of Information and Communication Technology*, Springer Nature Switzerland, pp. 274-285, doi: 10.1007/978-3-032-01429-0_24.

Aylsworth, T. and Castro, C. (2024), "Should I use ChatGPT to write my papers?", *Philosophy and Technology*, Vol. 37 No. 4, p. 117, doi: 10.1007/s13347-024-00809-w.

Bhaskar, P., Khan, M. and Bhaskar, P. (2025), "Exploring educators' perspectives on ChatGPT integration in engineering and management education", *Journal of Information, Communication and Ethics in Society*, Vol. 23 No. 4, pp. 526-549, doi: 10.1108/JICES-04-2025-0088.

Bisquerra Alzina, R. and Pérez-Escoda, N. (2015), "Les escales de Likert poden augmentar en sensibilitat?", *REIRE Revista D'Innovació I Recerca En Educació*, Vol. 8 No. 2, pp. 129-147, doi: 10.1344/reire2015.8.2828.

Bulut, O., Beiting-Parrish, M., Casabianca, J.M., Slater, S.C., Jiao, H., Song, D., Ormerod, C., Fabiyi, D.G., Ivan, R., Walsh, C., Rios, O., Wilson, J., Yildirim-Erbasli, S.N., Wongvorachan, T., Liu, J.X., Tan, B. and Morilova, P. (2024), "The rise of artificial intelligence in educational measurement: opportunities and ethical challenges", *Chinese/English Journal of Educational Measurement and Evaluation*, Vol. 5 No. 3, doi: 10.59863/MIQL7785.

Chen, F.F. (2007), "Sensitivity of goodness of fit indexes to lack of measurement invariance", *Structural Equation Modeling: A Multidisciplinary Journal*, Vol. 14 No. 3, pp. 464-504, doi: 10.1080/10705510701301834.

Downloaded from www.emerald.com/jices/article-pdf/doi/10.1108/JICES-01-2026-0003/11689032/jices-01-2026-0003en.pdf by guest on 09 November 2026

Chen, H.-J. (2025), "Verifying the link of innovativeness to the confirmation-expectation model of ChatGPT of students in learning", *Journal of Information, Communication and Ethics in Society*, Vol. 23 No. 3, pp. 433-447, doi: 10.1108/JICES-07-2024-0086.

Chen, S.-W. (2023), "Learning motivations and effort beliefs in confucian cultural context: a dual-mode theoretical framework of achievement goal", *Frontiers in Psychology*, Vol. 14, doi: 10.3389/fpsyg.2023.1058456.

Coldwell, W. (2024), "I received a first but it felt tainted and undeserved': inside the university AI cheating crisis", available at: www.theguardian.com/technology/2024/dec/15/i-received-a-first-but-it-felt-tainted-and-undeserved-inside-the-university-ai-cheating-crisis

Correa, D., López-Díez, J. and Campuzano-Hoyos, J. (2024), "A humanistic manager? Integrating humanities into professional management education in Colombia: the case of EAFIT university, 1994–2007", *Journal of Management History*, Vol. 31 No. 1, pp. 154-173, doi: 10.1108/JMH-10-2023-0110.

Dahl, B.M., Vasset, F. and Frilund, M. (2023), "Students' approaches to scientific essay writing as an educational method in higher education: a mixed methods study", *Social Sciences and Humanities Open*, Vol. 7 No. 1, p. 100389.

De Andrés-Sánchez, J. (2025), "The role of ethical perceptions in the acceptance of innovative sport technologies for competition: analysis of carbon plate running shoes", *Cultura, Ciencia y Deporte*.

De Ruyter, D. and Schinkel, A. (2017), "Ethics education at the university: from teaching an ethics module to education for the good life", *Bordon: Revista de Pedagogía*, Vol. 69 No. 4, pp. 125-138, doi: 10.13042/Bordon.2017.690409.

Dul, J. (2016), "Necessary condition analysis (NCA): logic and methodology of "necessary but not sufficient" causality", *Organizational Research Methods*, Vol. 19 No. 1, pp. 10-52, doi: 10.1177/1094428115584000.

Dul, J. (2024a), "A different causal perspective with necessary condition analysis", *Journal of Business Research*, Vol. 177, p. 114618, doi: 10.1016/j.jbusres.2024.114618.

Dul, J. (2024b), "How to sample in necessary condition analysis (NCA)", *European J. of International Management*, Vol. 23 No. 1, pp. 1-12, doi: 10.1504/EJIM.2024.138446.

Farinosi, M. and Melchior, C. (2025), "I use ChatGPT, but should I?' a multi-method analysis of students' practices and attitudes towards AI in higher education", *European Journal of Education*, Vol. 60 No. 2, p. e70094, doi: 10.1111/ejed.70094.

Faul, F., Erdfelder, E., Buchner, A. and Lang, A.-G. (2009), "Statistical power analyses using G*power 3.1: tests for correlation and regression analyses", *Behavior Research Methods*, Vol. 41 No. 4, pp. 1149-1160, doi: 10.3758/BRM.41.4.1149.

Gorrieri, L. (2025), "Should I use ChatGPT as an academic aid? A response to aylsworth and castro", *Philosophy and Technology*, Vol. 38 No. 1, p. 8, doi: 10.1007/s13347-024-00831-y.

Gosepath, S. (2021), "Equality", In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*, (Summer 2021 ed.). Metaphysics Research Lab, Stanford University, available at: https://plato.stanford.edu/archives/sum2021/entries/equality/.com

Han, B., Nawaz, S., Buchanan, G. and McKay, D. (2025), "Students' perceptions: exploring the interplay of ethical and pedagogical impacts for adopting AI in higher education", *International Journal of Artificial Intelligence in Education*, Vol. 35 No. 4, doi: 10.1007/s40593-024-00456-4.

Hasanein, A.M. and Sobaih, A.E.E. (2023), "Drivers and consequences of ChatGPT use in higher education: key stakeholder perspectives", *European Journal of Investigation in Health, Psychology and Education*, Vol. 13 No. 11, pp. 2599-2614, doi: 10.3390/ejihpe13110181.

Hua, J. and Cunningham, J. (2025), "Two years after ChatGPT: a thematic analysis of first-year students' reflections on AI tool use in higher education", *Journal of Information, Communication and Ethics in Society*, doi: 10.1108/JICES-05-2025-0118.

Downloaded from www.emerald.com/jices/article-pdf/doi/10.1108/JICES-01-2026-0003/11689032/jices-01-2026-0003en.pdf by guest on 09 November 2026

Jung, I. (2009), "Ethical judgments and behaviors: applying a multidimensional ethics scale to measuring ICT ethics of college students", *Computers and Education*, Vol. 53 No. 3, pp. 940-949, doi: 10.1016/j.compedu.2009.05.011.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F. and Kasneci, G. (2023), "ChatGPT for good? On opportunities and challenges of large language models for education", *Learning and Individual Differences*, Vol. 103, p. 102274, doi: 10.1016/j.lindif.2023.102274.

Killen, M. (2018), "The origins of morality: social equality, fairness, and justice", *Philosophical Psychology*, Vol. 31 No. 5, pp. 767-803, doi: 10.1080/09515089.2018.1486612.

Leonard, L.N., Riemenschneider, C.K. and Manly, T.S. (2017), "Ethical behavioral intention in an academic setting: models and predictors", *Journal of Academic Ethics*, Vol. 15 No. 2, pp. 141-166, doi: 10.1007/s10805-017-9273-2.

Li, X., Zhou, J. and Cheng, Y. (2025), "Bridging AI ethics between communication and computer science: a care ethics approach to foster organization-public relationships", *Journal of Information, Communication and Ethics in Society*, doi: 10.1108/JICES-03-2025-0069.

Licht, K. D F. (2024), "Generative artificial intelligence in higher education: why the "banning approach" to student use is sometimes morally justified", *Philosophy and Technology*, Vol. 37 No. 3, p. 113, doi: 10.1007/s13347-024-00799-9.

Maslow, A.H. (1966), *The Psychology of Science: A Reconnaissance*, Harper and Row.

Meyer, J.G., Urbanowicz, R.J., Martin, P.C., O'Connor, K., Li, R., Peng, P.C. and Moore, J.H. (2023), "ChatGPT and large language models in academia: opportunities and challenges", *BioData Mining*, Vol. 16 No. 1, p. 20, doi: 10.1186/s13040-023-00329-7.

Miles, L., Vaz, S. and Hinton, T. (2022), "Academic integrity in a time of pandemic: the impact of remote learning and assessment on student honesty", *Journal of University Teaching and Learning Practice*, Vol. 19 No. 4, pp. 79-95, doi: 10.53761/1.19.4.6.

Pelegrín-Borondo, J., Arias-Oliva, M., Murata, K. and Souto-Romero, M. (2020), "Does ethical judgment determine the decision to become a cyborg?", *Journal of Business Ethics*, Vol. 161 No. 1, pp. 5-17, doi: 10.1007/s10551-018-3970-7.

Pérez-Portabella, A., Arias-Oliva, M., Padilla-Castillo, G. and de Andrés-Sánchez, J. (2025), "Passing with ChatGPT? Ethical evaluations of generative AI use in higher education", *Digital*, Vol. 5 No. 3, doi: 10.3390/digital5030033.

Perry, A.H., Rettinger, D.A., Stephens, J.M., Anderman, E.M., McTernan, M.L., Tatum, H., McNally, D., Cullen, C. and Bertram Gallant, T. (2025), "From institutional climate to moral attitudes: examining theoretical models of academic misconduct", *Ethics and Behavior*, Vol. 36 No. 3, pp. 1-18, doi: 10.1080/10508422.2025.2514577.

Prashar, P., Gupta, Y. and Dwivedi, Y.K. (2023), "Plagiarism awareness efforts, students' ethical judgment and behaviors: a longitudinal experiment study on ethical nuances of plagiarism in higher education", *Studies in Higher Education*, Vol. 49 No. 6, pp. 929-955, doi: 10.1080/03075079.2023.2253835.

Reidenbach, R.E. and Robin, D.P. (1990), "Toward the development of a multidimensional scale for improving evaluations of business ethics", *Journal of Business Ethics*, Vol. 9 No. 8, pp. 639-653, doi: 10.1007/BF00383391.

Rosen, S. and Dreskin, S. (2024), "Cheating in school: a nuanced issue?", The Berkeley High Jacket, available at: https://berkeleyhighjacket.com/section/opinion/cheating-in-school-a-nuanced-issue

Sebastião, R., Kottekkadan, N.N., Thomas, T.K. and Niyas, M.K.K. (2025), "Generative AI tools (ChatGPT) in social science research", *Journal of Information, Communication and Ethics in Society*, Vol. 23 No. 2, pp. 284-290, doi: 10.1108/JICES-10-2024-0145.

Downloaded from www.emerald.com/jices/article-pdf/doi/10.1108/JICES-01-2026-0003/11689032/jices-01-2026-0003en.pdf by guest on 09 November 2026

Shawver, T.J. and Sennetti, J.T. (2009), "Measuring ethical sensitivity and evaluation", *Journal of Business Ethics*, Vol. 88 No. 4, pp. 663-678, doi: 10.1007/s10551-008-9973-z.

Sinnott-Armstrong, W. (2023), *Consequentialism, the Stanford Encyclopedia of Philosophy*, (WinterEditionWinterEdition), available at: https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=consequentialism

Strzelecki, A. (2024), "To use or not to use ChatGPT in higher education? A study of students' acceptance and use of technology", *Interactive Learning Environments*, Vol. 32 No. 9, pp. 5142-5155, doi: 10.1080/10494820.2023.2209881.

Suler, J. (2004), "The online disinhibition effect", *CyberPsychology and Behavior*, Vol. 7 No. 3, pp. 321-326, doi: 10.1089/1094931041291295.

Tlais, S., Alkhatib, A., Hamdan, R., HajjHussein, H., Hallal, K. and Malti, W.E. (2025), "Artificial intelligence in higher education: early perspectives from lebanese STEM faculty", *TechTrends*, Vol. 69 No. 3, pp. 1-9, doi: 10.1007/s11528-025-01061-5.

Venkatesh, V. and Davis, F.D. (2000), "A theoretical extension of the technology acceptance model: four longitudinal field studies", *Management Science*, Vol. 46 No. 2, doi: 10.1287/mnsc.46.2.186.11926.

Yang, S.C. (2012), "Ethical academic judgments and behaviors: applying a multidimensional ethics scale to measure the ethical academic behavior of graduate students", *Ethics and Behavior*, Vol. 22 No. 4, pp. 281-296, doi: 10.1080/10508422.2012.672907.

Zhao, L., Mao, H., Compton, B.J., Peng, J., Fu, G., Fang, F., Heyman, G.D. and Lee, K. (2022), "Academic dishonesty and its relations to peer cheating and culture: a meta-analysis of the perceived peer cheating effect", *Educational Research Review*, Vol. 36, p. 100455, doi: 10.1016/j.edurev.2022.100455.

Corresponding author
Jorge de Andrés-Sánchez can be contacted at: jorge.deandres@urv.cat

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: **permissions@emeraldinsight.com**

Downloaded from www.emerald.com/jices/article-pdf/doi/10.1108/JICES-01-2026-0003/11689032/jices-01-2026-0003en.pdf by guest on 09 November 2026