Revista Latinoamericana de Tecnología Educativa, 25(2)
https://doi.org/10.17398/1695-288X.25.2.201

RELATEC
Revista Latinoamericana de Tecnología Educativa

Received: December 19, 2025
Reviewed: April 20, 2026
Accepted: June 15, 2026

Corresponding authors' address:

[1] Departamento Economía de la Empresa. Facultad de Ciencias de la Economía y Empresa. Universidad Rey Juan Carlos. Paseo de los Artilleros, s/n, 28032 - Vicálvaro – Madrid (Spain)

[2] Facultad de Derecho, Empresa y Gobierno. Universidad Francisco de Vitoria. Ctra. Pozuelo-Majadahonda km. 1.800. 28223, Pozuelo de Alarcón - Madrid (Spain)

E-mail / ORCID

emilio.velasco@urjc.es

https://orcid.org/0000-0001-8765-9753

sofia.ruiz@ufv.es

https://orcid.org/0000-0002-7325-7777

ARTICLE / ARTÍCULO

Generative Artificial Intelligence, Perceived Effort, and Critical Thinking: Evidence from an Experimental Study with ChatGPT in Higher Education

Inteligencia artificial generativa, esfuerzo percibido y pensamiento crítico: evidencias de un estudio experimental con ChatGPT en Educación Superior

Emilio Velasco-Bartolome[1] & Sofía Ruiz-Campo[2]

Abstract: The emergence of Artificial Intelligence (AI) has had a profound impact on diverse sectors, including education. While the integration of technology into classrooms has been a constant for decades, both educators and students face new challenges due to the advent of NLP systems. Educators deal with analyzing and evaluating the originality of an essay. With the application of AI, students can answer theoretical questions and solve complex exercises, but grapple with the challenge of collaborating and engaging with a tool that relies on natural language instructions but can also provide outright incorrect responses. Our study aims to assess the effectiveness of ChatGPT in solving a case study with quantitative questions. We seek to compare the performance of students who use the tool with those who do not, from formulating instructions to validating or assessing the results. The data collected through pre-, and post-test questionnaires reveal consistently better results in students who did not use ChatGPT. However, the expectation of result accuracy and confidence in the outcomes obtained are similar in both groups. We suggest this to be interpreted as an effect that we call "Blind to Fraud" describing the situation in which students appear unaware of the likelihood of inaccuracy when using an AI-based NLP tool, which shows overconfidence in the results obtained. This can significantly impact learning outcomes and should be counterbalanced through critical thinking skills in education.

Keywords: Artificial Intelligence, ChatGPT, Illusion of Explanatory Depth, Critical Thinking, Higher Education.

Resumen: La aparición de la Inteligencia Artificial (IA) ha tenido un profundo impacto en diversos sectores, incluida la educación. Aunque la integración de la tecnología en las aulas ha sido una constante durante décadas, tanto el profesorado como el alumnado se enfrentan a nuevos desafíos debido a la irrupción de los sistemas de procesamiento del lenguaje natural (PLN). El profesorado debe afrontar el análisis y la evaluación de la originalidad de los ensayos. Mediante el uso de la IA, el alumnado puede responder preguntas teóricas y resolver ejercicios complejos, pero se enfrenta al desafío de colaborar e interactuar con una herramienta que se basa en instrucciones expresadas en lenguaje natural, pero que también puede proporcionar respuestas completamente incorrectas. Nuestro estudio tiene como objetivo evaluar la eficacia de ChatGPT para resolver un estudio de caso con preguntas cuantitativas. Pretendemos comparar el rendimiento del alumnado que utiliza la herramienta con el de quienes no la emplean, desde la formulación de instrucciones hasta la validación o evaluación de los resultados. Los datos recopilados mediante cuestionarios administrados antes y después de la prueba muestran resultados sistemáticamente mejores entre el alumnado que no utilizó ChatGPT. Sin embargo, las expectativas sobre la exactitud de los resultados y la confianza en los resultados obtenidos son similares en ambos grupos. Proponemos interpretar este fenómeno como un efecto que denominamos «ceguera ante el fraude», que describe una situación en la que el alumnado parece no ser consciente de la probabilidad de obtener resultados inexactos al utilizar una herramienta de PLN basada en IA, lo que se manifiesta en una confianza excesiva en los resultados obtenidos. Este fenómeno puede afectar significativamente a los resultados de aprendizaje y debería contrarrestarse mediante el desarrollo de habilidades de pensamiento crítico en la educación.

Palabras-clave: Inteligencia Artificial, ChatGPT, Ilusión de capacidad explicativa, Pensamiento crítico, Educación Superior.

1. Introduction

The widespread access to Natural Language Processing (NLP) models based on Artificial Intelligence (AI) has made a significant impact across various sectors (Odden et al., 2021), ranging from economics, business, and medicine to education. NLP models rapidly revolutionize teaching and learning by enabling personalized learning and on-demand support. Universities have long pursued pedagogical innovations, integrating technology to strengthen student-teacher relationships and equip students with real-world problem-solving skills. This paradigm includes technology-based learning, social networks, and global interactivity (Harasim (2000), Raja and Nagasubramani (2018)). Therefore, not only are innovative teaching techniques necessary but also their integration into the learning process.

Evidently, NLP models are here to stay, if not forever. Students are familiar with them, use them, and know their advantages. Evidence from a large-scale global study involving over 23,000 students across 109 countries confirms that, while students value ChatGPT for simplifying complex information and summarizing content, they find it considerably less reliable for supporting classroom learning (Ravšelj et al., 2025). Recent literature highlights that NLP-based AI systems such as ChatGPT simultaneously present substantial opportunities and significant pedagogical risks for higher education institutions (Fuchs, 2023). A systematic review of ChatGPT's limitations across multiple fields identifies accuracy and reliability concerns, constraints on critical thinking and problem-solving, and ethical and privacy issues as the most prominent challenges associated with its use (Sok & Heng, 2024). These findings are particularly relevant in educational contexts, where students' inability to detect or verify incorrect outputs may have direct consequences on learning outcomes. Academic institutions can either attempt to prevent their use, which requires significant efforts focused on plagiarism. Alternatively, academic institutions and educators can accept their use and adapt teaching methodologies to show the benefits of AI without students adopting the passive role of copying and pasting responses. Among other tasks, educators must prepare and adapt their classes and activities for AI use so that AI becomes a complement and collaborator rather than a substitute for students' (and teachers') work. This approach can enhance learning by involving students, as they need to acquire new skills, such as effectively expressing questions, objectives, and instructions in natural language, as well as validating or evaluating the obtained responses.

Among the alternatives available, ChatGPT is the most popular NLP model. Developed by OpenAI, it utilizes a vast dataset to generate text responses to user queries, feedback, and prompts (Gilson et al., 2023). Therefore, considering the trend in the university towards the use of ICT (Information and Communication Technology) in the classroom, this study empirically analyzes whether the use of ChatGPT can, as suggested by the

literature, complement the learning process, and enhance students' knowledge.

The study was conducted using an experimental methodology with international and Spanish students. The research aims to assess the performance, effort, and confidence of students using ChatGPT to solve a quantitative case study. Beyond its capacity for data processing, we operated under the assumption that ChatGPT could offer strategic value to solving the case study. Based on our initial assumptions regarding its utility, we highlighted these specific tasks: clear identification of business problems; identification of key data to solve a problem; identification of the necessary analytical techniques to solve a problem using the provided data; calculation and numerical results obtained by applying the necessary techniques; analysis of the validity of the obtained data; relevance of the data in relation to the problem; contextualized interpretation of the results; providing appropriate solutions by by providing informed insights.

The results are evaluated and discussed considering the following presumptions, which are based on Lo (2023), our experience, and the literature on students' competency outcomes in higher education (Nikolaos et al., 2019; Liesa-Orus et al., 2020). They are presented in Table 1.

Table 1. Presumptions of the research establish that «the group that uses ChatGPT vs non ChatGPT users…»

Id.	Presumptions
P.a	Achieves a higher grade in the case study.
P.b	Performs better in identifying company needs and goals.
P.c	Has a greater level of accuracy while using the appropriate technique for knowledge needs.
P.d	Has a lower level of accuracy while calculating the outcomes.
P.e	Has a greater level of accuracy when analysing statistical validity.
P.f	Has a higher accuracy in interpreting and suggesting solutions based on the data gathered.
P.g	Has a greater rate of response and consequently is able to answer more questions.
P.h	Has higher grade expectations.
P.i	Perceives a lower effort invested to solve the case study.
P.j	Would exhibit lower confidence due to their inability to oversee or verify the results.

Given the appeal of digital tools to young people (Irfan et al., 2023), the study explores students' perceptions of utilizing the tool and the consequences of its use on learning as well.

The study's findings can be useful for educational institutions and educators since they quantify the benefits of using AI-based NLP vs not using it.

It also investigates students' perspectives after completing a task utilizing this technology. Furthermore, the appropriate use of ChatGPT also implies a set of skills that must be developed during the university period, particularly two: the ability to ask in an orderly and conducive manner to a result (identification of task objectives), and the ability to interpret the results obtained: coherence, problem-solving ability, and reliability, among others.

In this context, we introduce the concept of "blindness to fraud" to describe the situation in which students appear unaware of the likelihood of inaccuracy when using an AI-based NLP tool and hence lack awareness of their genuine degree of understanding in the field. This dynamic could facilitate the replication of this study in multiple subjects with different core abilities involved, as well as giving a framework for contemplation on how to design tasks that can be performed in 'partnership' with AI.

Finally, teachers and institutions can investigate what complementary learning is required (communication with AI, learning to prompting, computational thinking, and skills to verify the results obtained) for the effective integration of AI into the classroom, i.e., to accelerate and personalize learning for university students, particularly in avoiding the "Blind to Fraud" effect. The article continues with the review of the literature, analyzes the methodology, followed by the results and conclusions of the study.

1.1. Review of Literature

The application of digital tools in education has broadened since Vygotsky and Cole's (1978) Activity Theories. Individuals, according to their viewpoint, exist inside a sociocultural framework and constantly shape their surroundings, interacting and developing new cultural mechanisms. According to Leontiev (1978), interaction is more than just a reaction; it causes internal changes that promote one's own development. Furthermore, Westin et al. (2015) investigated the role of ICT in their interactions with students and teachers, and how this improves interactivity and learning.

However, not all examples of technology adoption in the classroom provide favorable results (Moss et al., 2007), as educational outcomes largely depend on teachers' pedagogical approaches and the ways technology is integrated into learning processes. In this context, the emergence of new digital technologies does not diminish the role of educators; rather, it transforms their responsibilities toward guiding students in the effective and critical use of technological tools (Rawlins & Kehrwald, 2010). Teachers must therefore adapt their methodologies to ensure that AI-based tools maximize not only access to content, but also the development of meaningful learning processes and analytical skills.

From this perspective, AI holds substantial promise for improving learning, teaching, and support practices in higher education (Smith, 2022). For

example, AI-powered chatbots can provide immediate support to students by enhancing their learning experience, motivation, and engagement (O'Connor and McAndrew, 2021). Educators face a dual challenge: on the one hand, teachers must adapt their method to AI (Firat, 2023), and on the other, they must develop evaluation methods that enhance the detection of assignments whose authorship can be substantially attributable to AI. Simultaneously, students confront the problem of 'collaborating' with AI to improve their abilities, which necessitates skills in natural language querying and evaluating responses not only for correctness but also for relevance to the inquiry's goals.

Since ChatGPT's public release, scholars have examined its implications for higher education from multiple perspectives, including opportunities for productivity and learning, risks for assessment and academic integrity, ethical concerns, and disciplinary applications. Examples of numerous publications on this subject include Arif et al. (2023); Farrokhnia et al. (2023); Fauzi et al. (2023); Gilson et al. (2023); Mhlanga (2023); Rospigliosi (2023); Rudolph et al. (2023); Thurzo et al. (2023); Tlili et al. (2023). Tajik and Tajik (2023) add a further layer, categorizing the impacts of ChatGPT, or the deployment of other NLP-based AI models, into three categories: students, teachers, and the educational institution itself.

According to recent studies, the employment of AI in higher education can result in significant improvements across a variety of areas. AI-based educational tools have been associated with enhanced learner engagement, personalized learning experiences, and improved academic support (Yang et al., 2020; Tate et al., 2023). In particular, Nisar and Aslam (2023) found that ChatGPT can function as an effective self-learning and rapid-reference tool for students studying pharmacology, highlighting its potential to support autonomous learning processes. Similarly, Firat (2023) argues that ChatGPT may transform self-directed and open education environments by encouraging students to take greater responsibility for their own learning outcomes.

The study emphasizes students' need to master formulating relevant questions, engaging constructively with the NLP tool, and evaluating its responses. This involves overcoming challenges in framing queries, synthesizing information, and validating data to solve complex case studies effectively.

Another key difficulty that students may face is the possibility of being overly reliant on technology, which can undermine the development of critical thinking abilities such as the capacity to evaluate source quality and dependability, make informed decisions, and produce new ideas. Students may fall victim to what Rozenblit and Keil (2002) refer to as «the illusion of explanatory depth», in which they assume they have a thorough understanding of a subject despite having only superficial knowledge (Chromik, 2021). It is worth emphasizing that, to the best of our knowledge, this effect has been most prominently found in tasks demanding explanatory

knowledge and complicated causal patterns, but not in other activities such as following a method or providing factual descriptions. Chromik's (2021) article empirically establishes the reality of this impact.

According to Lo (2023), ChatGPT performs remarkably well in economics, satisfactorily in programming, and disappointingly in fields such as mathematics. Regardless of the areas where it demonstrates higher proficiency, this performance variation is exciting for a thorough case study, as presented to students in the current experiment, where performance may be examined in a variety of areas, including statement interpretation, computation, and proposal writing, all based on the case study findings.

2. Methodology

This paper describes an exploratory experimental investigation in which results were examined using inferential statistics.

2.1. The experiment

A total of 30 students were divided into two groups with the goal of resolving the same case study: 1) a control group (50% of students) that will use the statistical software (SPSS) learned during the subject, and 2) another group that will exclusively use ChatGPT. All students had already completed the entire curriculum (Market Research) and had experience working with similar case studies using SPSS.

The task required students to solve a case study consisting of thirteen questions with the goal of assessing the three distinct skill categories that they ought to have mastered at the end of the course. For the competency rubric, the questions are grouped according to the skill category involved and the competency area, according to the following scheme shown in Table 2.

Table 2. The case study examined three skill categories, raised thirteen implicit questions about the skills, and evaluated three competency groups.

Categories of skills evaluated in the study		**Questions implied in every skill**	**Groups of competencies assessed through the questions**
Category #1: Recognition of business problems, needs and goals that can be solved through quantitative market research		1	Critical reading, understanding of a business context
Category #2: Identification and application of statistical analysis techniques to solve	2.1. Statistical Techniques to be applied	4, 7 & 10	Statistical techniques, calculation, and validity analysis

problems. Divided into 3 components:	2.2. Results obtained in the analysis	2, 5, 8 & 11	
	2.3. Validity of the results	6, 9 & 12	
Category #3: Interpretation of outcomes considering the problems and needs of the business represented in the case study		3 & 13	Selection and interpretation of outcomes, as well as proposed business solutions

2.2. Construct and instruments' validation

For this exploratory research, a convenience sampling method was employed due to the accessibility of subjects within the academic environment. To strengthen internal validity, a split-classroom design was used to conduct a small-scale Randomized Controlled Trial (RCT).

Participants from the same class were randomly assigned to either the Chat GPT or SPSS group. This approach is methodologically more robust than using intact groups as it eliminates selection bias. Furthermore, since all students belong to the same section, they share identical prior training, which minimizes external variables and allows for a more precise comparative evaluation of the results in a natural classroom setting (ecological validity). This sampling approach was adapted from Verenikina's framework (2010) to the context of a public university in Madrid.

Since this research constitutes an exploratory experiment, our initial analysis primarily relies on inferential statistics, which includes T-Test for independent groups, and the effect sizes attributable to the use of ChatGPT. Occasionally, when the effect size is small or to isolate specific variables, correlations are applied.

Along with the 13-question case study, every student was asked to complete two questionnaires (Tables 3 and 4). The first questionnaire was delivered before accessing the case study and gathers sociodemographic (questions 1-3) and academic data (questions 4-11). The second questionnaire gathers perceptions about the task (questions 12-17), after completing the case study but before knowing the results of the exercises. The questionnaire development process was conducted under the supervision of one of the co-authors of the research, a pedagogue in case study methodology, and reviewed by three experts.

Table 3. First questionnaire (before the task): academic and character variables*

Item	Questions

1	Age
2	Country
3	Educational background
4	What´s your perception about your current skills on using AI?
5	What is approximately your GPA (average of your previous marks) in your university degree)?
6	What is the average score you obtained in previous assignments in Market Research II?
7	According to your own estimation, how many sessions of Market Research II, expressed as a percentage, did you attend during the semester? (Consider 1 as less than 10%, 2 as 20%, and so on until 10 which represents 100%)
8	How strong do you feel with your current knowledge in Market Research II?
9	How strong do you feel is your current knowledge concerning the interpretation of client's business needs and business goals?
10	How strong do you feel is your current knowledge concerning the different statistical techniques applied to analyse data in market research?
11	How strong do you feel is your current knowledge concerning the interpretation of results to provide suggestions or solutions based on data analysis results?

* Scale 1-10

The second questionnaire (Table 4) includes students' estimation of the outcomes achieved after completing the activity. Only for students who use ChatGPT does it incorporate additional questions, the number 13 and 17.

Table 4. Second questionnaire (after task): students' estimation of the outcomes achieved*

Item	Questions
12	What is the mark you consider you obtained in the case study?
13	How easy was to find the way (language, structure, procedures...) to cooperate with ChatGPT?
14	Regarding the effort required to solve the case study… how easy was it?
15	How confident do you feel about your answers?
16	Consider that this would have been a professional environment and the case study a real case of a real client. How confident would you feel about presenting these results to a client?
17	If you had to distribute the mark in the case study between the AI and you, what would be your share?

* Scale 1-10

The fact that we compare data given by students makes it easier to compare the results produced using the two alternative data gathering approaches. On the one hand, we inquire using numerical rating scales (ranging from 1 to 10) that correspond to both questions about their perceived

knowledge of the subject as a whole and their level of competence in each of the critical skills involved in solving the case study (questions 1 to 11); Cronbach's Alpha yields a result of 0.885, indicating an excellent level of reliability for the construct.

In contrast, Likert-type scales were used to assess the level of agreement with a proposition. This scale is appropriate for establishing comparisons when alternatives are difficult to assess because they are dependent on individuals' proclaimed perceptions. This type of scale is used to assess students' confidence in the results obtained, on a scale of 1 to 10, with 1 indicating "No confidence at all" and 10 indicating "Total confidence" (Questions 12 to 17). Cronbach's Alpha is applied again. The result is outstanding, demonstrating a high level of construct reliability (0.943).

Finally, when we use Cronbach's Alpha to test the degree of the overall construct reliability, this time with all the questions and scales employed. We get a result of 0.800, indicating that the scales and the construct itself have statistical validity.

3. Results

Once the suitability of the measurement construct was confirmed, our primary goal prior to analysis was to evaluate the groups' comparability. There is no important difference in the information collected from questions 1, 2 and 3, since these are students aged 20-21, a high percentage from Europe, and with similar previous studies.

Regarding questions 4 to 11 (Table 3), the results obtained in each group (Table 5) are also very similar. The result has a high statistical significance ($p < 0.05$), which means that it is statistically legitimate.

Table 5. Information gathered for the two student groups from the initial questionnaire. Questions 4-11*

Questions #	ChatGPT Users	Non ChatGPT Users
4: Current skills using AI	7.33	7.14
5: Current GPA	7.27	7.15
6: Previous marks in Market Research II	6.87	7.14
7: Attendance during the semester	7.13	7.07
8: Confidence in knowledge of Market Research II	5.73	5.00
9: Confidence in business needs and goals	6.00	5.64
10: Confidence in statistical techniques	5.20	4.79
11: Confidence in interpretation of results and solutions	5.80	5.14

* Scale 1-10

Because there are no discernible differences between the two student groups, we proceeded with the analysis..

3.1. Performance in competencies in the subject and performance in the task [Pre-Assumptions P.a) to P.f)]

Following that, we sought to investigate the Presumptions presented in Table 2. The summary can be found in Table 6, where a column has been added that shows the grade obtained in the case study.

Table 6. Actual results and standard deviation on the case study divided into categories* for both groups of students.

Category of skills	ChatGPT Users	StDev (%)	Non ChatGPT Users	StDev (%)
1. Recognition of Business needs and goals	1.72	0.11	3.85	0.08
2.1. Statistical techniques to be applied	3.75	0.05	6.63	0.05
2.2. Results obtained in the analysis	0.66	0.17	5.90	0.06
2.3. Validity of the results	5.00	0.06	5.64	0.06
3.Interpretation of the results	2.17	0.05	5.95	0.04
Grade achieved in the case study	1.72	0.69	4.91	0.61

* Scale 1-10

3.2. Presumption P.a): Grade of the case study

Starting with Presumption P.a), that establishes that "the group that uses ChatGPT achieves a higher grade in the case study", the data unequivocally demonstrate that ChatGPT users performed noticeably worse on average than non-users, while having a relatively high self-perception of competence with AI tools (Table 5). However, the average might conceal a high data dispersion, which was verified by calculating the standard deviation.

The standard deviation is low and very similar in both groups of students (except for the competency of calculating results for the group that used ChatGPT). Therefore, the average can be considered consistent for the analysis. This allows us to reject our assumption P.a), although a further analysis will be conducted to determine the possible cause.

3.3. Presumptions P.b) to P.f): Performance and accuracy.

The effect of the previous mark on the subject.

As the performance difference is so significant, we wanted to rule out any relationship between GPA or current subject grade (Question 5) with the outcome. The correlation result is statistically valid ($p<0.05$) only for the

relationship between the Previous Mark in the subject and skill #1. "Identification of Business needs" which shows a positive moderate intensity (+0.435) and the skill #4. "Validity of the results" (+0.431). Therefore, both skills could be explained for the whole group by the influence of the current mark in the subject. The remaining variables have a significance greater than 0.05 and are thus dismissed.

To complete the exploration of these skills, a test of Compared Means for independent samples was conducted since the most plausible explanation was that the use or non-use of ChatGPT could account for these differences. The result obtained is statistically valid (sig. <0.05), but it shows low to moderate effect sizes (Table 7) (in this case, Cohen's D is chosen because the standard deviations were similar).

Table 7. Cohen's D to measure the effect size in skills 2.1, 2.2 and 3.

Skills	(2.1) Statistical techniques to be applied	(2.2) Results obtained	(3) Interpretation of results obtained
Cohen's D (points stimation)	0.28	0.34	0.17

Therefore, the use of ChatGPT cannot be statistically attributed to a greater or lesser demonstration of performance in the skills described. Thus, we could challenge the conclusion of Lo (2023), which qualitatively summarizes perceptions from different studies about ChatGPT's performance in various competency areas. According to our data, the direct influence of ChatGPT usage on the performance in the measured competencies in isolation cannot be attributed because prior subject knowledge has a more significant influence. This allows us to reject, either through a correlation analysis or through the effect size, our assumptions P.b) to P.f) except P.c), because ChatGPT users do not achieve a higher grade in the case study (P.b); do not perform better in identifying company needs and goals (P.d); do not have a greater level of accuracy when analysing statistical validity (P.e) neither have a higher accuracy in interpreting and suggesting solutions based on the data gathered (P.f). Concerning P.c) we can confirm the presumption because ChatGPT users show lower level of accuracy while calculating the outcomes.

It can also be added, in relation to P.a) that when considering the final grade in the case study for both groups of students (and not just the result in each competency), the difference can be explained by the effect size, as in this case, the result of Cohen's D is 2.24. In other words, the final grade in the study significantly depends on the use or non-use of ChatGPT for its resolution. In this case, students who used ChatGPT exhibited significantly lower performance. Therefore, the use of ChatGPT negatively affects the overall case study

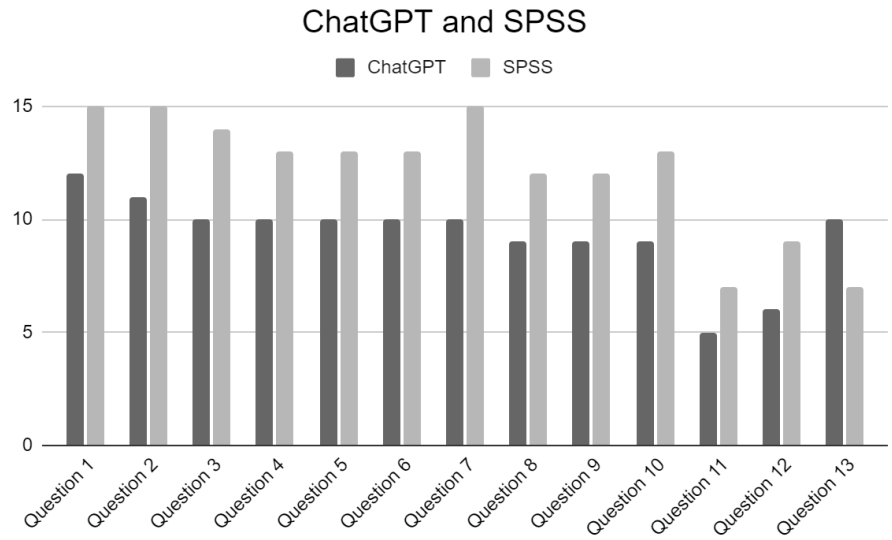
resolution process. Since the final grade depends on the partial grades, and statistical significance for individual competencies showed that the result in every skill was not attributable to the use of ChatGPT, although the result was clearly lower, it is possible to conclude that the focus should not be directed toward different skills, as concluded by Lo (2023) in their review study, but in the main task where the contribution of ChatGPT is decisive and reduce significantly the students' performance.

In short, without detracting from the individual competencies involved, we seek a holistic impact on the very process of developing solutions to complex tasks.

3.4. Presumption P.g): Rate of response

Another noteworthy aspect of performance is the rate of response of both groups of students. Figure 1 provides a summary of the number of responses per question in the case study.

Figure 1. Number of students responding per case study question, using ChatGPT and SPSS (Non ChatGPT users).



In all competencies (see Table 2) except for Category#3: Interpretation of results (question 13) ChatGPT users answered fewer questions, which contradicts our assumption P.g): "The group that uses ChatGPT has a greater rate of response and consequently is able to answer more questions".

It is significant, nonetheless, that 10 out of the 15 students who used ChatGPT answered the last question, but their result was notably lower (2.17/10) compared to only 7 students in the group that did not use ChatGPT,

who obtained (5.95/10). They were required to develop a buyer persona using the information from the case study. Students who used ChatGPT provided an answer, although in most cases, it was not based on the case study data or often related to the company's situation. However, students trusted in a well-structured, narrated, and plausible response provided by ChatGPT.

Therefore, the use of ChatGPT did not result in a higher response rate overall, implying greater real ease in completing the case study, so assumption P.g) must be rejected.

3.5. Presumption P.h): Grade expectations

After the resolution of the case study, not only were students graded, but they were also asked about the grade they expected to receive. While students who used ChatGPT expected better grades in the case study compared to those who used SPSS, their actual grades were significantly lower than anticipated. Conversely, students who did not use ChatGPT slightly underestimated their scores. Table 8 illustrates the disparities between estimated and actual grades for both groups.

Table 8. Actual score and estimated score for both groups of students, after the case study resolution.

	Actual	Estimated
ChatGPT Users	1.72	5.00
Non ChatGPT Users	4.91	4.31

When we apply the independent samples T-test to the grade estimation variable, we obtain a Cohen's D value of 2.52, indicating that the use or non-use of ChatGPT clearly influences grade estimation. Hence, although ChatGPT usage impacts academic performance by lowering it, it does raise grade expectations significantly, resulting in a noticeable overestimation of expected grades. This allows us to confirm our assumptions P.h) that establishes that the group that uses ChatGPT has better grade expectations.

According to the data analysis, we can conclude that the use of ChatGPT did not lead to higher performance in task resolution quality, nor in the grades obtained, nor in the number of questions they were able to answer; but surprisingly did so for the performance's expectation. Subsequent sections of this article will explore the potential reasons and implications of this overestimation.

3.6. Presumption P.i): Effort required

Apart from the grade, we were interested in the effort required as ChatGPT should contribute to reducing it. Comparing ChatGPT users and non-user students concerning the perception of effort invested to solve the case study, students using ChatGPT perceived a lower effort (Table 9).

Table 9. Perceived effort to solve the case study for both groups of students.

	ChatGPT Users	Non ChatGPT Users
Effort	5.5	7.15
StDev	3.23	1.67
StDev%	58.69%	23.4%

Given the significantly higher standard deviation among students who used ChatGPT, an analysis was conducted to assess if prior subject knowledge or overall GPA explained this variation. The result of a correlation analysis between both GPA and Subject Mark with the effort required is not statistically significant. Considering that the difference is high and cannot be explained by a confounding variable, we decided to conduct a T-Means test for independent groups.

The effect size measurement using Glass's Delta revealed a statistically significant result, with a large effect size (Glass's Delta = 3.228), indicating that the use of ChatGPT had a substantial impact on perceived differences in effort. The study also examined how the perception of effort influenced the estimation of obtained grades for both groups. The results, which were statistically valid (sig < 0.05), demonstrated a moderate (-0.4) and negative influence.

The negative, moderate, and statistically valid correlation between effort and grade perception implies that when less effort was perceived, students tended to expect higher grades. This may contribute to the greater optimism observed among students who used ChatGPT, as they perceived their effort to be significantly lower.

One potential factor influencing the perceived effort of students who used ChatGPT is the difficulty in using the tool itself, rather than the difficulty of the case study. The study investigated this by asking students to rate the ease of cooperating with ChatGPT on a scale from 0 (Easy) to 10 (Hard). The result was 3.71, indicating a moderate/low level of difficulty in working with the tool. In Figure 2, we show the greater difficulties in a frequency table.

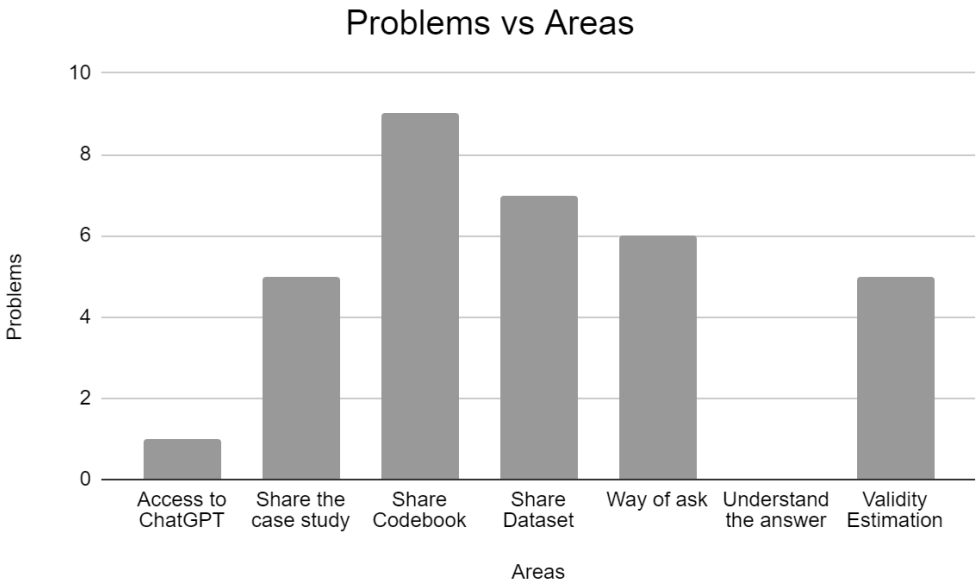


Figure 2. Frequency of problems found when interacting with ChatGPT by different areas.

In terms of effort needed to complete the study, students using ChatGPT had an average of 46.9 interactions with the NLP, almost one interaction per minute, with a high standard deviation of 0.62 Variation Coefficient.

An analysis of the correlation between the effort required to solve the case and the ease of cooperation with the tool shows a highly influential negative correlation (-0.57) between them, indicating that the greater the difficulty in using or cooperating with the AI, the greater the perceived effort. It does not seem risky to assume, therefore, that much of the perceived effort was invested in using the tool. This suggests that the effort dedicated to the case was perceived as very low or that the focus was more on using the tool and less on the case study, with students accepting the answers uncritically once obtained. The fact that no student reported difficulties in understanding the answers clearly confirms this.

Furthermore, as we showed previously, there was a significant difference in the number of interactions needed within students using ChatGPT. To analyze this high standard deviation, a Pearson correlation coefficient was calculated with both GPA and the current grade in the subject. The result showed no significant correlation with GPA, but it was significant with the current grade in the subject, yielding a result of -0.522 (two-tailed significance 0.06). In other words, students who performed better in the subject needed significantly fewer interactions with ChatGPT. Higher subject grades correlate with fewer interactions needed with the NLP tool.

Distribution of workload and marks between the student and AI

Then, we analysed students' credit allocation between themselves and ChatGPT. When asked to distribute marks on a scale from 0 to 10, where 0 implied the AI did all the work, and 10 implied the student did everything with the AI as a simple calculator, the average response was 3.82, indicating that students believed they contributed slightly less than the AI to solve the case (less than 50%).

The study further investigated what influenced this distribution. We considered the ease of cooperation with the AI and found a negative correlation of -0.675 (with statistical significance). Then, the more difficult it was to cooperate with the AI, the less students perceived the AI's contribution. This reinforced the idea that students' perceived effort primarily stemmed from the difficulty of using the NLP tool, rather than the case study itself.

Additionally, there was a moderate negative correlation (-0.593) statistically valid between the grade estimation and the distribution of contribution. In other words, the higher the grade estimation, the less contribution students attributed to the AI.

The study consistently found that:

a) Students' effort depended more on ChatGPT's usability than task difficulty.

b) Contribution distribution was strongly influenced by the ease of cooperating with the AI.

c) Higher self-assessed performance led to reduced attribution of achievement to AI.

3.7. Presumption P.j): Evidence of confidence in the results

Students were informed about using ChatGPT, but when it came to an answer as simple as a mean, 60% gave incorrect answers and 40% skipped basic calculations, neglecting simpler tools like calculators. In contrast, students who did not have access to ChatGPT answered correctly 93.3% of the time. Students trusted ChatGPT's answers without verification, claimed credit for solutions, and reported no difficulty understanding its responses.

The study also investigated the level of confidence both student groups had in their outcomes (Table 10). Students were questioned regarding their level of confidence in both academic and real-world professional settings. There were no differences between the two groups, although as could be predicted, students felt more confident in an academic atmosphere where less responsibility was suggested than in a professional one.

Table 10. Confidence expressed in the results obtained in the case study.

	Confidence in the academic field	**Confidence in a professional environment**
ChatGPT users	4.21	3.54
Non ChatGPT users	4.31	3.62

Contrary to the assumption that GPA or subject grade correlates with confidence, the results fail to support this, as the statistical significance exceeds 0.05. The T-test for independent groups does not offer any clarity on this matter either. Therefore, the initial assumption P.j): "Students who use ChatGPT would exhibit lower confidence in the results due to their inability to oversee or verify them", cannot be confirmed or rejected.

As we expected, higher grade expectations lead to greater confidence for both academic (0.893) and professional (0.809) purposes, and these correlations are statistically valid in both groups.

The results highlight two perceptions:

a) Accuracy of results: ChatGPT users exhibit high self-confidence, often overestimating the accuracy of their results due to a "blindness to fraud," despite a mismatch with actual performance.

b) Translation of results: This inflated confidence has little impact on their perceived utility or confidence in academic and professional applications.

To conclude the analysis of the results obtained in the study, a table is presented that summarizes the confirmation or not of the assumptions (Table 11).

Table 11. Summary of assumption rejection or confirmation and employed technique parameters.

	Statistical technique	**Status**	**Reason**
P.a)	T-means for independent groups	Rejected	Cohens' D > 2
P.b)	Correlation with GPA	Rejected	sig = +0.435. ChatGPT is a confounding variable.
P.c)	T-Test Independent samples	Rejected	Not attributable to ChatGPT Cohen's D = 0.28
P.d)	T-Test Independent samples	Rejected	Not attributable to ChatGPT Cohen's D = 0.34
P.e)	Correlation with GPA	Rejected	Sig = +0.431. ChatGPT is a confounding variable.
P.f)	T-Test Independent samples	Rejected	Not attributable to ChatGPT Cohen's D = 0.17

	Statistical technique	Status	Reason
P.g)	Chi-square	Rejected	Count of ChatGPT users is lower than count of SPSS users.
P.h)	T-Test	Confirmed	Cohens' D = 2.52
P.i)	T-Test	Confirmed	Glass' Delta = 3.228
P.j)	T-Test	Neither confirmed nor rejected	Sig > 0.05 in all tests

4. Conclusions

AI, particularly NLP models, is reshaping higher education, raising concerns about data security, performance, and legitimacy in educational settings. Educational institutions and their members bear the responsibility of ensuring that AI benefits both students and educators. In this sense, objective research is essential in decision-making rather than relying on subjective or persuasive expert's arguments about AI.

The study originally aimed to assess students' effort and outcomes when using AI, but it made an unexpected discovery concerning the confidence exhibited when using Natural Language Processing (NLP) AI.

The study presents two main areas of investigation. Firstly, it dissects the competencies required at a university level to examine them. The results indicate that the use of ChatGPT alone does not improve competency outcomes individually. Instead, students' prior knowledge of the subject appears to be a more explanatory factor for academic performance when considering each competency separately. However, when measuring overall test performance without breaking down competencies, ChatGPT usage negatively affects outcomes.

These results may stem from tool-related issues and from the ways students interact with AI, highlighting the need to train students in the effective and critical use of AI-based tools. In this regard, recent studies have emphasized that the educational value of ChatGPT depends largely on how students and educators harness their potential within structured learning processes (Zhu et al., 2023).

Although previous studies have highlighted ChatGPT's potential as a self-learning and academic support tool in higher education (Nisar & Aslam, 2023), our findings suggest that its effectiveness strongly depends on students' ability to critically evaluate and validate AI-generated outputs. In their study Aiyappa et al. (2023) argue that, despite their apparent usefulness evaluating ChatGPT and similar large language models reveal limitations associated with data contamination, as they are trained in the entire web. However, in our

investigation, the main problem appears to emerge at a later stage, beyond the fact-checking process itself (Irfan & Murray, 2023).

According to our findings, a more fundamental issue exists, one that can bemore directly addressed by educators, institutions, and students. Students tended to place an excessive degree of trust in responses perceived as personalized and authoritative, almost as if they were consulting a personal expert. This behavior is consistent with the phenomenon of cognitive offloading, whereby the delegation of cognitive tasks to external tools reduces individuals' engagement in independent reasoning and is associated with a measurable decline in critical thinking skills (Gerlich, 2025). While GPA may explain some competency-related outcomes, ChatGPT appears to have limited influence on the remaining abilities. Consequently, the presence of a conversational and seemingly tailored AI system may overshadow students' previously acquired analytical and critical-thinking skills.

The second area of study focuses on students' perceived effort. Students using ChatGPT reported lower effort levels, particularly those with better prior knowledge of the subject. Interestingly, the number of questions they asked ChatGPT did not affect their total response rate. Students who reported less effort and used ChatGPT had higher grade expectations and displayed greater confidence in their responses.

Thus, we began to suspect that there was a hidden effect in the use of the tool. AI tools with NLP models are designed to engage in conversations with users, conveying a sense of fluidity and ease in the exchange that may reinforce the perception of the validity of the results obtained. Students tend to consider that they don't get an answer gathered from the internet (as in a browser), but an answer created specifically by Intelligence just for them and for the occasion.

At this point, we explored the concept of Illusion of Explanatory Depth (Rozenblit and Klein, 2002) to attempt to explain our findings. According to the authors, individuals may believe they possess knowledge of the world simply because a certain social opinion state (from a group aligned with their beliefs) reinforces explanatory knowledge. Even if the person could verify the facts, this makes them think of something that is untrue overall. However, this heuristic did not entirely correspond to our findings, as it did not necessarily involve a social opinion state or an explanatory narrative.

The observed, and measured effect and its means (personable dialogue, low friction in responses, and individualized results) closely align with what Thaler and Sunstein termed a "Nudge" (Thaler & Sunstein, 2008). This refers to a choice architecture designed to guide our actions and perceptions beneath our rational radar.

Later, both Yeung (2017) and Lanzing (2019) ventured to suggest the category of "Hypernudging." In this model, choice architecture does not merely appeal to basic mechanisms shared by everyone; instead, it becomes individualized through algorithms and Big Data, creating a decision-making space that is intimate, personal, and unique.

Students prioritize their trust in a system that whispers in their ear like an expert over their own knowledge (even when they possess it). They do not think to consult other sources or verify information themselves, even though it would be very easy to do so. There is a "glow" in the interaction with AI—which would not be an exaggeration to call trust—that renders them "Blind to Fraud."

The effect we found was accurately observable in the final question of the case study, which concerned the construction of a buyer persona, where the response rate of students who used ChatGPT was high, but 100% of their responses were incorrect or completely misaligned with the case study. Although they already had the knowledge, they didn't even check whether the answer was true or not. In our opinion the problem is that the answer is plausible and tailored specifically for the case.

The observed phenomenon that we termed "Blindness to Fraud" appears to be unique to the use of AI tools in the context of NLP. It manifests as an initial perception of reduced effort, primarily attributed to tool usage rather than the specific task at hand. This reduced effort subsequently leads to inflated grade expectations, creating an unwarranted sense of self-attribution of achievement. Students who expected higher grades tended to attribute a significant share of the work to themselves when dividing it between themselves and the AI.

In contrast to the "Illusion of explanatory," the "Blindness to Fraud" phenomenon allows individuals to wholly attribute the answers or explanations to themselves, leading to a moral reflection that we shall set aside here for the sake of space. The private nature of AI interactions plays a significant role in this self-attribution, as it makes it easy to erase traces of external input. Understanding this phenomenon goes beyond teaching students to ask questions or verify results. It appears to be a bias originating from the interaction with AI systems delivered through NLP.

Therefore, we believe that a new subtask emerges for both students and educators, as well as institutions. Joyner (2023) suggests that alongside training in critical thinking and the capacity to challenge answers, there should be an emphasis on critically evaluating the processes that produce those answers. Utilizing ChatGPT effectively entails developing a specific skill set, particularly two key abilities that should be cultivated during the university experience: the skill to pose inquiries in a structured and conducive manner towards achieving objectives, and the capacity to interpret outcomes with coherence, problem-solving proficiency, and reliability, among other attributes.

5. References

Aiyappa, R., An, J., Kwak, H., & Ahn, Y. Y. (2023). Can we trust the evaluation on ChatGPT? *arXiv preprint arXiv:2303.12767*. https://doi.org/10.48550/arXiv.2303.12767

Arif, T. B., Munaf, U., & Ul-Haque, I. (2023). The future of medical education and research: Is ChatGPT a blessing or blight in disguise? *Medical Education Online*, *28*(1): 2181052. https://doi.org/10.1080/10872981.2023.2181052

Chromik, M., Eiband, M., Buchner, F., Krüger, A., & Butz, A. (2021*). I Think I Get Your Point, AI! The Illusion of Explanatory Depth in Explainable AI*. In 26th International Conference on Intelligent User Interfaces (IUI '21). Association for Computing Machinery, NY, USA, 307–317. https://doi.org/10.1145/3397481.3450644

Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2023). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*. https://doi.org/10.1080/14703297.2023.2195846

Fauzi, F., Tuhuteru, L., Sampe, F., Ausat, A. M. A., & Hatta, H. R. (2023). Analysing the role of ChatGPT in improving student productivity in higher education. *Journal on Education*, *5*(4), 14886-14891. https://doi.org/10.31004/joe.v5i4.2563

Firat, M. (2023). How ChatGPT Can Transform Autodidactic Experiences and Open Education? https://doi.org/10.31219/osf.io/9ge8m

Fuchs, K. (2023). Exploring the opportunities and challenges of NLP models in higher education: is ChatGPT a blessing or a curse? In *Frontiers in Education*, 8, 1166682. Frontiers. https://www.frontiersin.org/articles/10.3389/feduc.2023.1166682/full

Gerlich, M. (2025). AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. *Societies*, *15*(1), 6. https://doi.org/10.3390/soc15010006

Gilson, A., Safranek, C. W., Huang, T., Socrates, V., Chi, L., Taylor, R. A., & Chartash, D. (2023). How does ChatGPT perform on the United States medical licensing examination? The implications of large language models for medical education and knowledge assessment. *JMIR Medical Education*, 9(1), e45312. https://doi.org/10.3389/feduc.2023.1166682

Harasim, L. (2000). Shift happens: Online education as a new paradigm in learning. *The Internet and Higher Education*, 3(1-2), 41-61. https://doi.org/10.1016/S1096-7516(00)00032-4

Irfan, M., Murray, L., & Ali, S. (2023). Insights into Student Perceptions: Investigating Artificial Intelligence (AI) Tool Usability in Irish Higher Education at the University of Limerick. *Global Digital & Print Media Review*, VI(II), 48-63. http://dx.doi.org/10.31703/gdpmr.2023(VI-II).05

Irfan, M., & Murray, L. (2023). Empowering Critical skills of Teaching and Learning with Artificial Intelligence (AI) tools. University of Limerick. https://doi.org/10.34961/researchrepository-ul.22788506.v1

Joyner, D., (2023). *ChatGPT in Education: Partner or Pariah?* XRDS 29, 3 (Spring

2023), 48–51. https://doi.org/10.1145/3589651

Lanzing, M. (2019). "Strongly recommended" revisiting decisional privacy to judge hypernudging in self-tracking technologies. *Philosophy & Technology*, *32*(3), 549-568. https://doi.org/10.1007/s13347-018-0316-4

Leontiev, N., (1978). *Activity, consciousness and personality*. Prentice-Hall.

Liesa-Orus, M., Latorre-Cosculluela, C., Vázquez-Toledo, S., & Sierra-Sánchez, V. (2020). The technological challenge facing higher education professors: Perceptions of ICT tools for developing 21st century skills. *Sustainability*, 12(13), 5339. http://dx.doi.org/10.3390/su12135339

Lo, C. K. (2023). What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. Education Sciences, 13(4), 410. *MDPI* AG. http://dx.doi.org/10.3390/educsci13040410

Mhlanga, D. (2023). Open AI in education, the responsible and ethical use of ChatGPT towards lifelong learning. http://dx.doi.org/10.2139/SSRN.4354422

Moss, G, Jewitt, C, Levacic, R, Armstrong, V, Cardini, A. & Castle, F. (2007). The interactive whiteboards, pedagogy, and pupil performance evaluation: an evaluation of the Schools Whiteboard Expansion (SWE). Project London Challenge. https://www.researchgate.net/publication/267239102_The_Interactive_Whiteboards_Pedagogy_and_Pupil_Performance_Evaluation_An_Evaluation_of_the_Schools_Whiteboard_Expansion_SWE_Project_London_Challenge

Nikolaos, K., Tzafea, O., & Thanos, T. (2019). University students'skills in using the internet for educational purposes and the digital divide. *European Journal of Open Education and E-Learning Studies*, 4(1), 60-72. http://www.oapub.org/edu

Nisar, S., & Aslam, M. S. (2023). Is ChatGPT a Good Tool for T&CM Students in Studying Pharmacology? https://dx.doi.org/10.2139/ssrn.4324310

O'Connor, P. & McAndrew, S. (2021). Enhancing Learning Experiences with AI-Powered Chatbots in Higher Education. *Journal of Educational Technology, 48*(3), 167-185.

Odden, T. O. B., Marin, A., & Rudolph, J. L. (2021). How has Science Education changed over the last 100 years? *An analysis using natural language processing. Sci. Educ.* 105, 653–680. doi: 10.1002/sce.21623

Raja, R., & Nagasubramani, P. C. (2018). Impact of modern technology in education. Journal of Applied and *Advanced Research, 3*(1), 33-35. https://doi.org/10.21839/jaar.2018.v3iS1.165

Ravšelj, D., Keržič, D., Tomaževič, N., Umek, L., Brezovar, N., Iahad, N. A., ... & Aristovnik, A. (2025). Higher education students' perceptions of ChatGPT: A global study of early reactions. *PLoS One*, *20*(2), e0315011

Rawlins, P., & Kehrwald, B. (2010). Education technology in teacher education: Overcoming challenges, realizing opportunities. In *Cases on digital technologies in higher education: Issues and challenges* (50-63). IGI Global. https://doi.org/10.4018/978-1-61520-869-2.ch004

Rospigliosi, P. (2023) Artificial intelligence in teaching and learning: What questions should we ask of ChatGPT? *Interactive Learning Environments, 31*(1), 1–3.

https://doi.org/10.1080/10494820.2023.2180191

Rozenblit, L. & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive scie*nce, 26(5), 521-562. https://doi.org/10.1207/s15516709cog2605_1

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching, 6*(1). https://doi.org/10.37074/jalt.2023.6.1.9

Smith, J. (2022). Personalized Learning through AI in Higher Education. *International Journal of Educational Technology, 39*(2), 79-94.

Sok, S., & Heng, K. (2024). Opportunities, challenges, and strategies for using ChatGPT in higher education: A literature review. *Journal of Digital Educational Technology, 4*(1), ep2401. https://doi.org/10.30935/jdet/13152

Tajik, E. & Tajik, F. (2023, April). A Comprehensive Examination of the Potential Application of ChatGPT In Higher Education Institutions. *TechRxiv.* https://doi.org/10.36227/techrxiv.22589497

Tate, T., Doroudi, S., Ritchie, D. & Xu, Y. (2023). Educational Research and AI-Generated Writing: Confronting the Coming Tsunami. https://doi.org/10.35542/osf.io/4mec3

Thurzo, A., Strunga, M., Urban, R., Surovkova, J., & Afrashtehfar, K. I. (2023). Impact of artificial intelligence on dental education: a review and guide for curriculum update. *Education Sciences*, 13(2), 150. https://doi.org/10.3390/educsci13020150

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

Verenikina, I. (2010). Vygotsky in twenty-first-century research. In EdMedia+ innovate learning (pp. 16-25). Association for the Advancement of Computing in Education (AACE). https://ro.uow.edu.au/cgi/viewcontent.cgi?article=2337&context=edupapers

Vygotsky, L. S., & Cole, M. (1978). Mind in society: Development of higher psychological processes. Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

Westin, T., Wiklund, M., Mozelius, P., & Norberg, L. (2015). Regression From Game-Oriented to Traditional School. *Journal of Educational Technology Systems, 43*(4), 349–370. https://doi.org/10.1177/0047239515588

Yang, D., Oh, E. S., & Wang, Y. (2020). Hybrid physical education teaching and curriculum design based on a voice interactive artificial intelligence educational robot. *Sustainability*, 12(19), 8000. https://doi.org/10.3390/su12198000

Yeung, K., (2017) 'Hypernudge': Big Data as a mode of regulation by design, Information, Communication & Society, 20:1, 118-136, DOI: 10.1080/1369118X.2016.1186713

Zhu, C., Sun, M., Luo, J., Li, T., & Wang, M. (2023). How to harness the potential of ChatGPT in education? Knowledge Management & eLearning, 15(2), 133–152. https://doi.org/10.34105/j.kmel.2023.15.008