EDU**XX1**

FOURCHAT: intervention aimed at promoting the critical use of ChatGPT in university

FOURCHAT: intervención orientada a promover el uso crítico de ChatGPT en la universidad

Felipe Bravo-Duarte [1,2*] (iD)
Esther Villajos [2] (iD)
Aida Soriano [2] (iD)

[1] Facultad de Psicología, Universidad UNIACC, Chile
[2] IDOCAL-Universitat de València, Spain
* Corresponding Author. E-mail: felipe.bravo@uv.es

ABSTRACT

This study evaluates the effectiveness of FOURCHAT (FOrmación Universitaria en uso Responsable de CHATgpt), an educational intervention designed to promote the critical and responsible use of ChatGPT among university students. The research, which included 150 students of three different degrees (111 in the experimental group and 39 in the control group), analyzed changes in trust towards ChatGPT results and information verification behaviors, also examining the relationship of these variables with digital responsibility and critical thinking. For this, pre- and post-intervention measures were performed, comparing both groups through factorial ANOVAS and linear regressions. The results show that the intervention was effective, especially in increasing verification behaviors in the experimental group, while these decreased in the control group. Likewise, digital responsibility showed a significant relationship with reduced trust and increased post-intervention verification behaviors. In conclusion, this work highlights the importance of training students in the ethical use of artificial intelligence tools to favor their academic development and promote optimal inclusion of these technologies in higher education.

Keywords: artificial intelligence, ethics of technology, digitalization, responsibility (education), critical thinking, higher education

RESUMEN

Este estudio evalúa la efectividad de FOURCHAT (FOrmación Universitaria en uso Responsable de CHATgpt), una intervención educativa diseñada para promover el uso crítico y responsable de ChatGPT entre estudiantes universitarios. La investigación, que contó con 150 estudiantes de tres grados diferentes (111 en grupo experimental y 39 en control), analizó los cambios en la confianza hacia los resultados de ChatGPT y las conductas de comprobación de la información, examinando además la relación de estas variables con la responsabilidad digital y el pensamiento crítico. Para esto, se realizaron medidas pre y post intervención, comparando ambos grupos a través de ANOVAS factoriales y regresiones lineales. Los resultados muestran que la intervención fue efectiva, especialmente en el incremento de conductas de comprobación en el grupo experimental, mientras que estas disminuyeron en el grupo control. Asimismo, la responsabilidad digital mostró una relación significativa con la disminución de la confianza y el aumento de las conductas de comprobación post-intervención. En conclusión, este trabajo destaca la importancia de formar al estudiantado en el uso ético de las herramientas de inteligencia artificial para favorecer su desarrollo académico y fomentar una inclusión óptima de estas tecnologías en la educación superior.

Palabras clave: inteligencia artificial, ética de la tecnología, digitalización, responsabilidad (educación), pensamiento crítico, educación superior

INTRODUCTION

In recent years, the unstoppable advance of Artificial Intelligence (AI) has revolutionized multiple sectors, from medicine to entertainment. However, its most significant influence has been felt in the academic field, where AI has emerged as a transformative tool (Zawacki-Richter et al., 2019). This revolution has intensified with the arrival of generative AI tools, such as ChatGPT (OpenAI, 2023).

On one hand, ChatGPT offers significant benefits, such as the ability to analyze and synthesize large volumes of information (Zawacki-Richter et al., 2019) or provide quick and, generally, coherent responses to user queries (OpenAI, 2023). However, it also presents important limitations, such as the difficulty in establishing effective filters and restrictions (Bond et al., 2024), the existence of significant biases in training data (Lim et al., 2023), or the tendency to generate false or invented information, including non-existent citations, authors, or facts (Bond et al., 2024).

This has generated concern in the educational community, which faces the challenge of integrating these tools appropriately and ethically into teaching-learning processes. Unfortunately, in many cases, improper use of AI in academic work has been observed, manifesting in practices such as plagiarism or text manipulation to try to avoid detection (de Frutos et al., 2024; Sallam, 2023; Segarra et al., 2024). Although work is being done on regulations to make explicit when AI is used in an academic text, and there are tools that can help detect when a text has been generated or modified with AI, these have significant limitations, and, moreover, the development of new applications capable of evading these controls also advances at a dizzying pace (Grace et al., 2023).

Faced with this situation, some institutions have opted to establish protocols or even prohibitions in an attempt to regulate their use, measures that require time for implementation and that are not always effective given the accelerated pace with which AI is incorporated into our daily lives (García-Peñalvo et al., 2024); therefore, it has been pointed out that prohibiting or developing restrictive protocols will hardly manage to regulate the responsible use of AI or guarantee its adequate use (García-Peñalvo, 2023).

Considering all of this, it is evident that a possible viable and effective solution is to accept its use and integrate it in the best possible way in the academic context (García-Peñalvo et al., 2024; López et al., 2023). This would mean that, instead of adopting a defensive or rejecting stance, educational institutions would assume a proactive and constructive approach, focused on training the academic community for ethical, responsible, and critical use of AI (Floridi et al., 2018; Murtaza et al., 2022). Implementing this would imply, on one hand, providing faculty with the training and resources necessary to integrate these tools into their teaching practice effectively and pedagogically (Ocaña-Fernández et al., 2019); and, on the

other hand, it would involve implementing training programs aimed at developing students' digital competencies, critical thinking, and ethical awareness, harnessing the potential of AI while being aware of its limitations and risks (Bond et al., 2024; Vega et al., 2023). Only from this perspective, based on realistic acceptance and responsible integration of AI in the academic context, could it be possible to successfully face the challenges posed by this technological transformation and turn it into an opportunity to enrich and improve educational processes (UNESCO, 2021; Chen et al., 2020).

It is in this framework where the importance of the educational community facing the challenge of integrating generative AI in teaching-learning processes emerges. This integration must be both in the cognitive aspects of learning (that is, related to the mental processes of students; Sternberg & Sternberg, 1996), as well as in the behavioral ones (which refer to observable behaviors of students; Schunk, 2012). Cognitive processes in students involve the acquisition, processing, and application of information, while behavioral processes are manifested in active participation, interaction with others, and the performance of tasks (Zimmerman and Schunk, 2011). Therefore, it is important that this integration of AI is both in how students think, as well as in how they act (Mayer, 2002).

Additionally, systematic reviews such as that of Batista et al. (2024) regarding the integration of AI in academic work indicate the need to generate clear guidelines to ensure compliance with academic standards and promote conscious use of these tools. This need is supported by research such as that of Wang et al. (2024) in 104 American universities, which reveals a growing trend towards the implementation of pedagogical strategies specifically designed to develop students' critical thinking. Empirical evidence suggests, therefore, that the integration of generative AI in academic contexts should be based on two essential pillars: digital responsibility and the development of critical thinking. These pillars will be detailed below.

Digital Responsibility

Digital responsibility refers to the capacity to use digital tools consciously, taking into account the ethical and legal implications of online activity (Jara and Ochoa, 2020). This concept is closely related to digital ethics, which refers to the set of principles and values that should guide the behavior of individuals in the digital environment, promoting responsible, respectful, and conscious use of technologies (Floridi et al., 2018).

Currently, the European Union is legislating about the ethical and responsible use of AI, seeking to make systems and platforms transparent with their results, including the disclosure of when content is generated by AI and the prevention of illegal and/or copyrighted content (Madiega, 2024). However, much of the ethical

and responsible use remains in the hands of the end user, who must internalize that AI should be used consciously.

In the specific context of using ChatGPT, this implies avoiding practices such as plagiarism or improper appropriation of texts through the simple action of "copy and paste" (Jara and Ochoa, 2020). Likewise, it entails assuming an active and committed role in the management of AI tools, which translates into the need to properly identify and cite the sources of information used by the model to generate its responses (Baidoo-Anu and Ansah, 2023).

This practice is not only fundamental for recognizing the work of the original authors and avoiding plagiarism, but it also allows the user to assess the quality and reliability of the sources used by the AI, a key aspect for determining the confidence that can be placed in the results obtained (Lim et al., 2023). Additionally, digital responsibility implies understanding that the output generated by ChatGPT is not a final product, but raw material that requires review, editing, and elaboration by the user, to adapt it to their specific needs and the academic context in which it is framed (Bond et al., 2024).

Critical Thinking

Critical thinking is another fundamental competence that acquires special relevance in the context of using AI tools. It is a mental process that involves analyzing, evaluating, and questioning information in a reflective and well-founded manner, to form reasoned judgments and make informed decisions (Dwyer et al., 2014). In the specific case of interaction with ChatGPT, it means adopting an active and inquiring stance towards the results generated by the tool, avoiding accepting them immediately or placing blind trust in them (Vega et al., 2023).

This implies, on one hand, being aware of the potential limitations and biases of AI, and understanding that, although it can offer coherent and apparently well-founded answers, these are not always complete, accurate, or impartial (Lim et al., 2023). On the other hand, critical thinking in relation to ChatGPT entails assuming the responsibility of verifying and contrasting the information obtained, resorting to reliable external sources and contrasting the data and arguments presented by the AI with prior knowledge and with other opinions or theories on the topic (Bond et al., 2024). This prior knowledge can come, for example, from sessions with teachers or from the user's own experience.

This attitude of constant questioning and checking is essential to develop a realistic trust in the tool, based on the understanding of its strengths and weaknesses (Baidoo-Anu and Ansah, 2023). Likewise, critical thinking involves reflecting on the intentionality and context in which AI responses are framed and assessing their adequacy and relevance in relation to the specific objectives and needs of the

user in each situation (Parra-Sánchez, 2022). In fact, interventions with students such as that of Liu and Wang (2024) and research such as that of Ruiz-Rojas et al. (2024) indicate that the use of generative AI tools even increases critical thinking behaviors in students, due to the need to evaluate and synthesize the large amount of information these tools provide.

FOURCHAT: Training in Digital Responsibility and Critical Thinking

As we can see, digital responsibility and critical thinking are essential competencies that every student must develop. In an educational context increasingly permeated by digital technologies, and, in particular, by generative AI tools like ChatGPT, the acquisition of these skills is fundamental to ensure an adequate, reflective, and responsible use of these resources (UNESCO, 2021). For this reason, it is very important that these learnings are developed preferably at the beginning of their academic training, as this will not only lay the foundations for them to be trained as competent and reliable professionals, but will also contribute to their integral development as citizens, preparing them to face the challenges of an increasingly digitized world (Murtaza et al., 2022).

Recent interventions in the academic world demonstrate a growing interest in integrating digital responsibility and critical thinking in higher education. For example, in Norway, Styve et al. (2024) implemented a framework to develop critical thinking practices intertwined with the use of generative AI in an introductory programming course, obtaining positive results in fostering students' critical awareness. In turn, in the United States, Wood and Moss (2024) integrated the responsible use of generative AI in master's subjects, finding that structured teaching of good use of these tools can improve the understanding of their ethical implications. In Nigeria, Yusuf et al. (2024) implemented a program to improve critical thinking skills in the synthesis behaviors of AI-generated texts, demonstrating its effectiveness in graduate students.

However, there is a need to document more intervention experiences that specifically integrate these two dimensions—digital responsibility and critical thinking—in the university context, especially in Spain. As a result of this, a pioneering training with university students has been developed, focused on developing digital responsibility and critical thinking regarding the use of ChatGPT: FOURCHAT (*FOrmación Universitaria en el uso Responsable de ChatGPT* - University Training in the Responsible Use of ChatGPT).

FOURCHAT is a theoretical-practical training, in which the students themselves can interact directly with ChatGPT and verify for themselves the need for critical and reflective human intervention in working with generative AI (Lim et al., 2023). Likewise, this training helps to demystify the technology and develop a grounded

relationship of trust with it, based on understanding its potentialities and limitations, and on assuming an active and reflective role in its use (Baidoo-Anu and Ansah, 2023).

To evaluate the effects of this intervention, two variables have been considered: the first, at the cognitive level, relates to the degree of trust that students place in the veracity of the texts that AI produces (that is, how much they think that ChatGPT results are credible). The second variable, at the behavioral level, refers to the action of verifying the results that ChatGPT offers them, contrasting with other sources available for consultation. In this way, it can be evaluated if the intervention has been beneficial and effective, both in the way students think and in the way they act towards AI.

Therefore, the general objective of this research work is to evaluate the impact of the FOURCHAT intervention on the critical use of ChatGPT by university students, analyzing the changes in trust placed in the results, information verification behaviors, and examining the relationship of digital responsibility and critical thinking with these behaviors. This general objective consists of the following specific objectives: (a) to analyze the impact of FOURCHAT training on the trust that students place in the results offered by ChatGPT; (b) to analyze the impact of FOURCHAT training on the verification behaviors that students perform on the information offered by ChatGPT; (c) to analyze the relationship between digital responsibility with the level of trust and verification of students in relation to the results offered by ChatGPT; (d) to analyze the relationship between critical thinking with the level of trust and verification of students in relation to the results offered by ChatGPT.

Likewise, the one-directional hypotheses proposed are: (H_1) After the intervention, the experimental group trusts ChatGPT results less; (H_2) After the intervention, the experimental group verifies ChatGPT results more; (H_3) Digital responsibility is associated, negatively, with trust in ChatGPT results ($H_{3.1}$), and positively with verification of results ($H_{3.2}$); (H_4) Critical thinking is associated, negatively, with trust in ChatGPT results ($H_{4.1}$), and positively with verification of results ($H_{4.2}$).

METHODS

Design

The FOURCHAT intervention program was born from an educational innovation project recognized and funded by the university. For its creation, implementation, and evaluation, both professors and doctoral students and undergraduate students have participated. The study methodology was mixed, as qualitative methodology

(through a focus group) and quantitative (questionnaire) were used following positivist and interpretative paradigms. Likewise, the design was pretest-posttest with a control group, which allowed to compare observed changes between groups and times.

Participants

The initial sample consisted of 221 undergraduate students. Of those 221, 168 were assigned to the FOURCHAT experimental group and 53 to the control group. The students come from first and second year of three degrees from the same university: psychology, labor relations, and tourism, although only in the groups where the responsible professors taught. The study population consisted of all the students where the involved professors taught (5 professors). The sampling was non-probabilistic, as the entire population was invited to participate; however, not all students attended class because participation was voluntary. It should be noted that, given that the variables of interest of this research study focus on the use that students make or made of ChatGPT (e.g., to what degree they verify the results they obtain through ChatGPT), a filter question about previous use of ChatGPT was added to the pretest questionnaire (i.e., "Have you ever used ChatGPT?"). Those students who had not previously used it were not considered for this research.

Therefore, the final sample consisted of a total of 150 students (111 from the experimental group and 39 from the control group). Regarding demographic characteristics, the mean age of the experimental group was 19.05 years, with a standard deviation of 1.87. This group was composed of 21 people of male gender, 89 of female gender, and one non-binary person. On the other hand, the mean age of the control group was 21.33 years, with a standard deviation of 1.75. This group was composed of 13 people of male gender and 26 of female gender.

Procedure

The intervention is composed of the following phases:

Phase 1 - Focus group on the use of ChatGPT: During the first phase (2023-2024 academic year), all students were invited to participate in a focus group on ChatGPT. The call was made without providing additional information to avoid biases in the responses, and five female students volunteered to participate. Analysis of the focus group discussions revealed three main axes:

First, participants expressed concern about how ChatGPT could encourage avoidance of cognitive effort and deterioration of fundamental skills. As one

participant noted: "*[...] it's creating lazy people. 20 years ago, people read that 20-40 page report, now they don't [...] we're going to become very lazy*" (Student D). This concern is reinforced by another observation from the participants: "*it's very easy to succumb to laziness. Oh well, if I copy in this work it doesn't matter, in the next one I'll...*" (Student D).

Second, a particular concern emerged about the use of ChatGPT in educational levels prior to university and its possible consequences. One student highlighted: "*if you don't learn at that moment, when you get to university you won't know how to do it either*" (Student B), referring specifically to basic skills such as making summaries or text commentaries. This point relates to the observation of the same participant, related to "*it's true that in high school people are using it much more because in the end they are much easier topics that don't have the complexity of a university work, and in the end the chat has more than enough to make you a high school work*" (Student B).

Finally, the participants agreed that prohibition is not the appropriate solution, advocating for a more proactive educational approach. As one of them expressed: "*instead of demonizing ChatGPT, what the educational field should do is take a step forward and teach children and adults who are also studying to make good use and use it as an ally*" (Student C). This perspective is reinforced by the pragmatic observation of another participant: "*the question is to assume that people are going to use it. Whether they say yes or no, they will use it. Since they are going to use it, at least make it useful for something*" (Student B) and complemented by a second participant: "it would be better 'learn to use it well', not 'don't use it'" (Student A).

This proposal from the participants emphasized the importance of implementing specialized training programs in higher education that not only promote the responsible use of ChatGPT, but also contribute to the development of digital, ethical, and critical thinking skills among students. Thus, although the initial objective of the focus group was to gather information about the use of ChatGPT, the results pointed to a clear path towards the need for structured training by the faculty.

Phase 2 - Development of training material and questionnaires: The next phase consisted of developing the training material and questionnaires to measure the effectiveness of the intervention. The training material included aspects such as a brief introduction to AI, advantages, disadvantages, and risks of AI, the development of good prompts, ethical and legal aspects, and many examples of biases and/or errors in ChatGPT results obtained from the experience of academics and students. This material was developed by two professors with previous experience on AI in the academic world and served to implement phase 5 of the project (see Figure 1)

Figure 1
Extract from the teaching material developed for FOURCHAT training

For the preparation of the questionnaires, scales were sought to measure the use made of ChatGPT in terms of its reliability, ethical use, verification of results, among others. Specific information in this regard can be found in the measures section.

Phase 3 - Teacher training: The next stage of the intervention began with a series of meetings with the faculty in charge of implementing the project. The objective of these sessions was to train them in the necessary competencies to deliver the intervention contemplated in the FOURCHAT program and to jointly resolve any possible doubts that might arise. Therefore, the two professors who developed the teaching material explained to the rest of the professors involved (3 professors) how to use the material developed, standardizing the training and ensuring that everyone was on the same page in the explanation to the students.

Phase 4 - Group preparation and pretest: The students were randomly divided into an experimental group and a control group from the groups where the 5 professors had teaching duties. The assignment between the experimental group and the control group was done by lottery. There was no division within the groups, so all students in the same group (within their degree and subject) entered the control or experimental group. The experimental group carried out the entire FOURCHAT intervention, while the control group only carried out the pretest, a distractor activity, and the posttest. It should be noted that, for ethical-moral reasons and so that they would have the same learning opportunities as the rest of

their peers, once the entire intervention process was completed, the intervention for the experimental group was replicated for the control group.

Phase 5 - Expository training (class and expert talk): In this phase, only the students from the FOURCHAT experimental groups actively participated. First, a session was held where the professors presented the previously prepared information on the good use of ChatGPT. As already mentioned, this training was the same for all students in the experimental groups.

Subsequently, a talk was given with an AI expert external to the university, where students could learn in a more practical way the ethical implications that its misuse could have in the work world. In this space, they could reflect on the ethical aspects of managing AI tools and how this impacts their future employability. This talk was recorded with the aim of reproducing the recording for all experimental groups.

Phase 6 - Experiential activity with ChatGPT: In this phase, both the experimental and control groups participated, although with different instructions.

In the experimental group, an activity was carried out where, following the subject's theme, they could test ChatGPT. First, they had to ask for scientific references on some topic of the subject, to see if they really existed; second, they had to test ChatGPT's answer on a topic of their choice and verify to what extent it offered ethical responses or presented biases or errors. The idea was for them to interact with the tool in order to verify all the answers that ChatGPT offered to detect its flaws.

In the control group, a distractor activity was carried out, where they only were asked to search for academic information on a topic of interest.

Phase 7 - Posttest: The last phase of the intervention corresponded to the response to the posttest survey, both for the experimental group and the control group. The variables were the same as in the pretest, although some questions were added about the satisfaction of having participated in the project and about the perceived effectiveness of the training program.

Final results day: As a first result of the intervention, the research team organized a presentation day of results, where all students were invited to send their class work. The research team selected those that had higher quality and presented more interesting results. The day functioned as a scientific congress: there was a first part of a small presentation on AI in the university environment and, subsequently, the selected student groups presented their work orally.

All these phases were carried out during the first semester of the 23-24 academic year. Therefore, the intervention lasted three months (from October to December), with 5 hours divided into 7 different days throughout the semester.

Student participation in the entire process was voluntary and completely anonymous. The questionnaires (pre and post) were answered through the

LimeSurvey platform, online. During classes, students had the link or QR available to access the questionnaire and fill it out at the time.

Measures

The measures used in the study were as follows:

a) *Digital responsibility and critical thinking*: to measure these two variables, the questionnaire on the development of digital and socio-civic competence (DIGISOC) (Peart 2020; 2022) was used. Specifically, the dimensions of digital responsibility (5 items) and critical thinking (4 items) were used. The scale was a 5-point Likert-type anchoring (1=not at all agree; 5=strongly agree). Some examples of items are: *"Before performing a digital activity, I usually think about the possible consequences"* (for digital responsibility) and *"I am a person who is critical of the information that reaches me"* (for critical thinking). In the validation carried out by Peart et al. (2020), the general scale presented high reliability (α=.904).

b) *Trust:* To measure trust in the answers provided by ChatGPT, the following item was used: *"To what degree do you trust the information you obtain through ChatGPT?".* It was measured with a Likert-type scale with 5 anchor points (1=never trust; 5=completely trust).

c) *Verification:* An item was used to measure the degree to which students verify the results that ChatGPT gives them. The item was *"Do you usually check the veracity of the information you obtain through ChatGPT?"* This item was measured with a 5-point Likert-type scale (1=never check; 5=always check).

Analysis

Different analyses were carried out with the IBM SPSS v.28 software. (IBM Corp., 2021). Central tendency descriptive statistics (means, standard deviations, kurtosis, and skewness), 2x2 factorial ANOVAs (time: pretest and posttest; group: experimental and control) were calculated and, finally, linear regression analyses were performed to evaluate the proposed hypotheses. These regressions separated by group and time allow us to examine whether the intervention strengthened the associations between the promoted competencies (digital responsibility and critical thinking) and specific behaviors (trust and verification), especially in the experimental group.

RESULTS

Descriptives

Table 1 shows the means of the different variables studied, distinguishing between the experimental and control groups, as well as between the pretest and posttest measures for each of the groups. Due to the importance of ensuring the adequacy of the data to a normal distribution, normality was evaluated considering the kurtosis and skewness values. These values, being within the range of -2 to +2, indicated a satisfactory fit of the data to a normal distribution (Ryu, 2011).

Table 1

Descriptives of the studied variables

| Variable	Time	Group	Mean	S.D	Skewness	Kurtosis
Trust	Pretest	Experimental	3.36	.79	-.52	.39
		Control	3.28	.80	-.15	-.56
	Postest	Experimental	3.01	.82	-.09	-.59
		Control	3.04	.76	-.61	.45
Verification	Pretest	Experimental	3.01	1.20	-.05	-.82
		Control	3.36	1.44	-.23	-1.30
	Postest	Experimental	3.16	1.11	-.10	-.48
		Control	3.02	1.20	-.04	-.91
Digital responsibility	Pretest	Experimental	4.13	.55	-.66	.11
		Control	3.91	.56	-.46	.27
	Postest	Experimental	4.13	.63	-.76	.27
		Control	3.96	.55	-.44	.29
Critical thinking	Pretest	Experimental	4.09	.59	-.59	.11
		Control	4.08	.61	-.39	.07
	Postest	Experimental	4.13	.58	-.57	.02
		Control	4.02	.63	-.63	.89

Results of the FOURCHAT intervention: changes in trust and verification

First, when analyzing the impact of FOURCHAT training on the degree of trust that students have in ChatGPT results, it can be seen that, after the intervention, it decreased in both groups. However, this decrease was more pronounced in the experimental group, where trust decreased by .32 points, while in the control group it decreased by only .24 points.

When comparing the two groups through ANOVA, the difference was not significant and the effect size was extremely small (F(1, 346) = .311, p = .29, η²p = .001), so hypothesis 1 must be rejected from a statistical point of view. Despite this, direct observation of the data suggests that the intervention implemented in the experimental group had a more notable impact on the decrease in trust compared to the control group. In Figure 2, it is highlighted that the slope of the line representing the experimental group is more pronounced than that of the control group, indicating a faster reduction in this first group.

Figure 2
*Pre and post trust levels according to group (experimental versus control)*
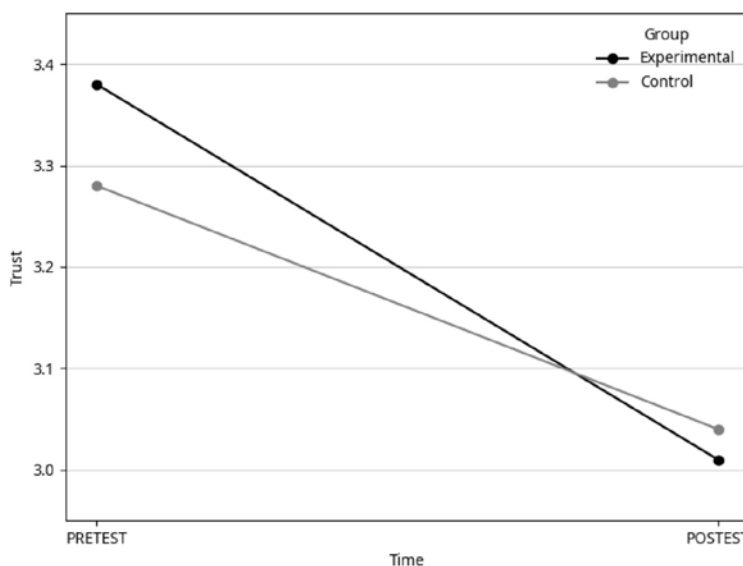
Regarding the effect of the FOURCHAT intervention on the level of verification of the responses obtained through ChatGPT, the results of the ANOVA support hypothesis 2. A significant difference was found in the changes between pre and posttest in the verification of results by students in the experimental group

compared to the control group ($F(1, 345) = 2.80$, $p < .05$ one-tailed), although the effect size was small ($\eta^2 p = .01$). The mean verification level of the experimental group increased from 3.01 in the pretest to 3.16 in the posttest. In contrast, the control group showed a decrease in verification of results, with a mean that fell from 3.36 in the pretest to 3.02 in the posttest. Therefore, while the rate of change was greater in the control group, this occurs in the inverse direction of the experimental group. The results are visualized in Figure 3.

Figure 3
Levels of pre- and post-testing by group (experimental versus control)
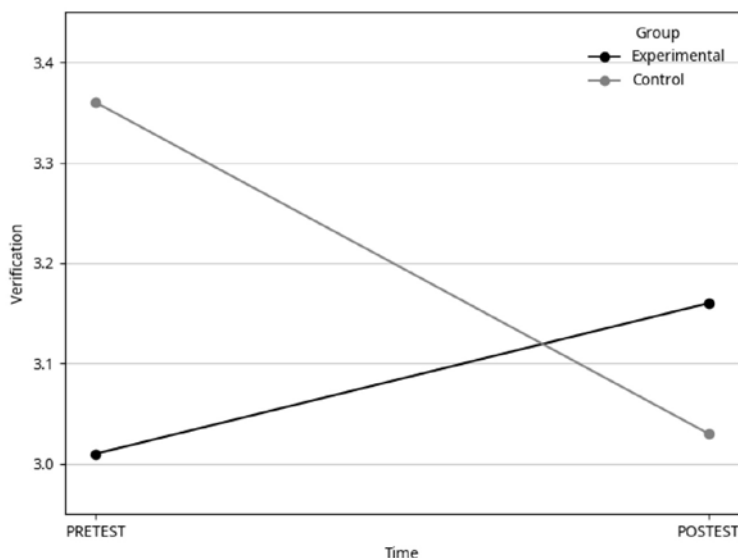
In summary, while the intervention could generate changes in both variables (trust and verification), FOURCHAT had a significant effect on the change in verification behaviors, where the experimental group showed an increase in these actions after the implementation of the program.

Results of regression analysis: digital responsibility and critical thinking as antecedents of trust and verification

As the next stage of our research, and to respond to our hypotheses 3 and 4, we sought to determine if the two variables that our FOURCHAT training promotes—digital responsibility and critical thinking—have a significant influence on trust in ChatGPT results and verification behaviors of students.

First, when analyzing the effect of digital responsibility and critical thinking on students' trust in ChatGPT, the regression analyses revealed a significant and negative relationship between digital responsibility and trust only in the experimental group during the posttest (see Table 2). In the pretest, neither digital responsibility nor critical thinking showed an association with levels of trust in either group; however, after the intervention, and in line with hypothesis 3.1, students in the experimental group who showed greater digital responsibility also showed less trust in ChatGPT responses (standardized β = -.23; p < .05 one-tailed), a relationship that was not observed in the control group. The emergence of this association exclusively in the experimental group after the intervention suggests that the FOURCHAT intervention, designed to develop digital responsibility, succeeded in activating the relationship between these variables. On the other hand, regarding hypothesis 4.1, no significant effects of critical thinking on trust were found in the posttest for any group.

Table 2

Regression of digital responsibility and critical thinking for trust

Group	Predictor	β	S.E	β (std)	Sig. (one-tailed)	β	S.E	β (std)	Sig. (one-tailed)
				Pretest				**Postest**	
Experimental	Digital responsibility	-.15	.14	-.11	.13	**-.30**	**.12**	**-.23**	**.01***
	Critical thinking	-.17	.13	-.13	.10	-.16	.13	-.11	.11
Control	Digital responsibility	-.03	.24	-.02	.45	-.18	.21	-.13	.19
	Critical thinking	-.01	.22	-.01	.50	.01	.18	.01	.47

* *p* < .05.

Regarding verification as a dependent variable (Table 3), the analyses also revealed differentiated results for each predictor. As with trust, digital responsibility showed a significant relationship with verification, only in the posttest of the experimental group (standardized β = .18; p < .05 one-tailed), supporting hypothesis 3.2 and suggesting that FOURCHAT activated the positive effect of this variable. On the other hand, in line with hypothesis 4.2, critical thinking was positively associated with verification of ChatGPT responses in the

experimental group, both in the pretest (standardized β = .21; p < .05 one-tailed) and in the posttest (standardized β = .20; p < .05 one-tailed). The similarity of these coefficients suggests that the FOURCHAT intervention only contributed to maintaining this relationship. On the other hand, none of these associations was significant in the control group.

Table 3
Regression of digital responsibility and critical thinking for verification

Group	Predictor	Pretest				Postest			
		β	S.E	β (std)	Sig. (one-tailed)	β	S.E	β (std)	Sig. (one-tailed)
Experimental	Digital responsibility	.32	.20	.15	.06	**.32**	**.16**	**.18**	**.02***
	Critical thinking	**.44**	**.19**	**.21**	**.01***	**.39**	**.17**	**.20**	**.01***
Control	Digital responsibility	-.10	.45	-.04	.41	.07	.33	.03	.41
	Critical thinking	.30	.41	.13	.27	.22	.28	.12	.22

* *p* < .05.

In summary, it could be seen that FOURCHAT generates greater changes in the digital responsibility of students, as it relates positively to the development of trust and verification, a phenomenon that is not visualized in the control group and that our intervention also enhances.

DISCUSSION

This research focused on four main objectives: 1) to analyze how FOURCHAT (FOrmación Universitaria en uso Responsable del CHATgpt) training affected students' trust in ChatGPT results; 2) to analyze whether this training influenced whether students verify the information they obtain from ChatGPT; 3) to analyze the relationship between digital responsibility and said degree of trust and verification of students; and 4) to analyze the relationship between critical thinking with the level of trust and verification of students in relation to the results offered by ChatGPT. In other words, our research aimed to verify whether teaching students about digital responsibility and critical thinking could help them use ChatGPT more adequately and reflectively.

The results demonstrate the effectiveness of the intervention, evidenced by the changes observed in the experimental group. Although very slight improvements were observed in the level of trust towards ChatGPT results, the most notable change occurred in the increase in verification behaviors among students who participated in FOURCHAT. This suggests that the training succeeded in developing greater critical awareness in the participants, motivating them to verify more frequently the answers obtained from ChatGPT. Likewise, these findings coincide with experiences documented in other countries, where teacher training oriented to the appropriate use of generative artificial intelligence has shown significant effects on the way students use these platforms within the academic context (Styve et al., 2024; Wood and Moss, 2024).

This contrast is especially significant when compared to the control group, where these behaviors even decreased during the second measurement. The decrease in verification in the control group could be explained by an unwanted adverse reaction: the fact that university professors dedicated time to question the use of ChatGPT through questionnaires could have activated alarm signals that led students to interpret that they should not use the tool, instead of learning to use it critically. By not receiving specific training on how to leverage the tool responsibly, they may have opted to reduce its use, which could decrease their verification behaviors.

On the other hand, we found that in students of the experimental group, digital responsibility was significantly related to both the decrease in trust in the tool and the increase in verification behaviors. This indicates that when students become aware of their active role regarding AI results and understand the origin of their training sources (Baidoo-Anu and Ansah, 2023), both cognitive changes (greater skepticism towards the veracity of the results) and behavioral changes (greater verification of the reliability of the sources) occur. However, a slight decrease was observed in the measure of digital responsibility during the posttest, possibly because students developed a more critical view of their own responsibility in the use of these tools.

Finally, regarding critical thinking, it does not show a significant relationship with trust towards the results produced by AI. In turn, it was observed that it already exerted an important positive influence on verification behaviors of the experimental group both before and after the intervention, unlike the control group. This could be explained by a previous development of critical thinking in the experimental group—possibly acquired in other subjects or previous educational experiences—that could enhance the effect of FOURCHAT training. Likewise, these results find support in studies such as that of Bond et al. (2024) and interventions such as that of Yusuf et al. (2024) that relate the promotion of critical thinking with a more conscious use of artificial intelligence.

Theoretical and practical implications

From our FOURCHAT intervention experience, a series of implications are derived, which are indicated below:

First, its implementation at the university level can be highly beneficial for students and the academic community in general. It is necessary to demonstrate that ChatGPT is a tool available to optimize academic activities and that it can be a valuable aid for their tasks, if it is used with responsibility and critical thinking. In addition, its scope could also be expanded to other educational levels.

Second, the integration of the promotion of digital responsibility and critical thinking in teaching processes is crucial to ensure an adequate use of ChatGPT and other AI tools. Only in this way can we transition from an approach centered on "pursuing the student to detect plagiarism by AI use," which can be considered a lost battle, to one oriented towards developing students' competencies in AI use for good use of this valuable tool. To achieve this, it is proposed to integrate the FOURCHAT intervention in the different aspects of learning, according to Bloom's taxonomy (Anderson and Krathwohl, 2001) with contributions from Mas et al. (2023):

– Remember: Students should keep in mind what ChatGPT consists of and how AI is fed. Likewise, they should identify which tool is the most appropriate to use within the wide range of available AIs (Mas et al., 2023).

– Understand: It is important that students understand that ChatGPT can yield erroneous or inaccurate results. For this, according to Mas et al. (2023), it is important that they learn to define precisely what is indicated in the prompts that are introduced into the tool.

– Apply: Students must know how to communicate with ChatGPT (for example, creation of prompts) for the responsible and ethical use of ChatGPT in different situations. For example, they could practice formulating clear and specific questions or instructions to obtain more precise and relevant answers from the tool.

– Analyze: It is necessary for students to develop skills to discriminate what information or results from ChatGPT really serve to do academic work. To do this, they could be provided with practical exercises in which they have to evaluate the quality and relevance of different responses generated by ChatGPT, contrasting them with other sources and justifying their assessments (Mas et al., 2023).

– Evaluate: Students must be able to evaluate the quality of the information that ChatGPT yields. An interesting activity would be to ask them to contrast the tool's responses with other reliable sources of information, identifying possible discrepancies or errors.

 – Create: Students must learn to generate their own texts from the ideas that ChatGPT can provide. To promote this, they could be proposed to use the tool as a starting point to develop an essay or report, incorporating their own reflections, arguments, and additional references. In this phase, they can also be urged to try new applications and uses of this tool, as well as combine instructions and results (Mas et al., 2023)

Finally, it is important to highlight the need to improve these initiatives through co-creative and participatory processes with students and technological referents, allowing to keep training updated in a constantly evolving field. This collaborative approach, which could be implemented through project-based learning (Bell, 2010), ensures a better understanding of student needs and emerging ethical challenges.

Limitations and future studies

The study presents some limitations that deserve consideration. The sample in the control group is relatively small compared to the experimental group, and the focus is exclusively on ChatGPT, although there are other relevant AI tools. Given its preliminary nature, efforts are being made to expand the sample, which could allow the formulation of new hypotheses and research objectives. Likewise, FOURCHAT training has the potential to be refined and applied in various educational levels and national and international academic institutions.

Conclusion

Based on the results obtained in our study, we conclude that the intervention carried out with university students is effective for them to make more responsible and ethical use of ChatGPT, increasing their critical thinking and information verification with respect to the results provided by it. This sharply affects the teaching-learning process, reaching higher levels of Bloom's taxonomy such as analyzing, evaluating, and creating. We believe that, by training university students in these matters, not only is their academic development favored, but a better inclusion of AI in higher education is also promoted.

In summary, this study makes a relevant and original contribution to knowledge about the responsible and ethical use of AI in the educational field, providing empirical evidence on the effectiveness of a pioneering training intervention in this field.

ACKNOWLEDGEMENTS

REFERENCES

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, *7*(1), 52-62. http://dx.doi.org/10.2139/ssrn.4337484

Batista, J., Mesquita, A., & Carnaz, G. (2024). Generative ai and higher education: trends, challenges, and future directions from a systematic literature review. *Information*, *15*(11), 676. https://doi.org/10.3390/info15110676

Bell, S. (2010). Project-based learning for the 21st century: Skills for the future. *The Clearing House*, *83*(2), 39-43.

Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: a call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education*, *21*(1), 4. https://doi.org/10.1186/s41239-023-00436-z

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, 8, 75264-75278. https://doi.org/10.1109/ACCESS.2020.2988510

De Frutos, N., Campo-Carrasco, L., Sainz de la Maza, M., & Extabe-Urbieta, J. M. (2024). Aplicación de la inteligencia artificial (IA) en educación: Los beneficios y limitaciones de la IA percibidos por el profesorado de educación primaria, educación secundaria y educación superior. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, *27*(1), 207-224. https://doi.org/10.6018/reifop.577211

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good ai society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*(4), 689-707. https://doi.org/10.1007/s11023-018-9482-5

García-Peñalvo, F. J. (2023). The perception of artificial intelligence in educational contexts after the launch of ChatGPT: Disruption or panic? *Education in the Knowledge Society, 24*, e31279. https://doi.org/10.14201/eks.31279

García-Peñalvo, F. J., Llorens-Largo, F., & Vidal, J. (2024). The new reality of education in the face of advances in generative artificial intelligence. [La nueva realidad de la educación ante los avances de la inteligencia artificial generativa]. *RIED-Revista Iberoamericana de Educación a Distancia*, *27*(1), 9-39. https://doi.org/10.5944/ried.27.1.37716

Grace, E. G., Vidhyavathi, P., & Malathi, P. (2023). A study on AI in education: opportunities and challenges for personalized learning. *Industrial Engineering Journal, 52*(05), 750-759. https://doi.org/10.36893/IEJ.2023.V52I05.750-759

IBM Corp. Released 2021. *IBM SPSS Statistics for Windows, Version 28.0.* Armonk, NY: IBM Corp

Jara, I., & Ochoa, J. M. (2020). *Usos y efectos de la inteligencia artificial en educación*. CAF - Banco de Desarrollo de América Latina. https://doi.org/10.18235/0002380

Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *International Journal of Management Education, 21*(2), 100790. https://doi.org/10.1016/j.ijme.2023.100790

Liu, W. & Wang, Y. (2024). The effects of using ai tools on critical thinking in english literature classes among efl learners: an intervention study. *European Journal of Education.* Education, 59, e12804 https://doi.org/10.1111/ejed.12804

López. A., Rodríguez, L & Montes, R. (2023). *Guía de uso de ChatGPT para potenciar el aprendizaje activo e interactivo en el aula universitaria.* Universidad Rey Juan Carlos. Recuperado de https://burjcdigital.urjc.es/handle/10115/22149

Madiega, T. (2024). *European Union Artificial Intelligence Act.* Retrieved from: https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf

Mayer, R. E. (2002). *Multimedia learning* (3rd ed.). Cambridge University Press.

Murtaza, M., Ahmed, Y., Shamsi, J. A., Sherwani, F., & Usman, M. (2022). AI-based personalized e-learning systems: Issues, challenges, and solutions. *IEEE Access*, *10*, 81323-81342. https://doi.org/10.1109/ACCESS.2022.3193938

Ocaña-Fernández, Y., Valenzuela-Fernández, L. A., & Garro-Aburto, L. L. (2019). Inteligencia artificial y sus implicaciones en la educación superior. *Propósitos y representaciones, 7*(2), 536-568. http://dx.doi.org/10.20511/pyr2019.v7n2.274

OpenAI. (2023). *GPT-4 Technical Report*. Retrieved from: https://doi.org/10.48550/arXiv.2303.08774

Parra-Sánchez, J. S. (2022). Potencialidades de la inteligencia artificial en educación superior: un enfoque desde la personalización. *Revista Tecnológica-Educativa Docentes 2.0*, *14*(1), 19-27. https://doi.org/10.37843/rted.v14i1.296

Peart, M. T., Gutiérrez-Esteban, P., & Cubo-Delgado, S. (2020). Development of the digital and socio-civic skills (DIGISOC) questionnaire. *Educational Technology Research and Development*, *68*, 3327-3351.

Peart, M. T. (2022). *El desarrollo de la competencia digital y socio-cívica para promover la educación para la ciudadanía digital y la participación juvenil* (Doctoral dissertation, Universidad de Extremadura).

Ruiz-Rojas, L. I., Salvador-Ullauri, L., & Acosta-Vargas, P. (2024). Collaborative working and critical thinking: adoption of generative artificial intelligence tools in higher education. *Sustainability, 16*(13), 5367. https://doi.org/10.3390/su16135367

Ryu, E. (2011). Effects of skewness and kurtosis on normal-theory based maximum likelihood test statistic in multilevel structural equation modeling. *Behavior research methods*, *43*, 1066-1074.

Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. *Healthcare, 11*(6), 887. https://doi.org/10.3390/healthcare11060887

Schunk, D. H. (2012). *Learning theories: An educational perspective* (8th ed.). Pearson.

Segarra, M., Grangel, R., & Belmonte O. (2024). ChatGPT como herramienta de apoyo al aprendizaje en la educación superior: una experiencia docente. *Tecnología, Ciencia y Educación*, *28*, 7-44. https://doi.org/10.51302/tce.2024.19083

Sternberg, R. J., & Sternberg, K. (1996). *Cognitive psychology* (7th ed.). Cengage Learning.

Styve, A., Virkki, O. T., & Naeem, U. (2024, mayo). Developing critical thinking practices interwoven with generative ai usage in an introductory programming course. *In 2024 IEEE Global Engineering Education Conference (EDUCON)* (pp. 01-08). IEEE. https://doi.org/10.1109/EDUCON60312.2024.10578746.

UNESCO (2021). *Inteligencia artificial y educación: Guía para las personas a cargo de formular políticas.* UNESCO. Retrieved from: https://unesdoc.unesco.org/ark:/48223/pf0000379376

Vega, J., Borja, E. E., & Ramírez, P. J. (2023). ChatGPT e inteligencia artificial: ¿obstáculo o ventaja para la educación médica superior? *Educación Médica Superior, 37*(2), e3851.

Wang, H., Dang, A., Wu, Z., & Mac, S. (2024). Generative ai in higher education: seeing chatgpt through universities' policies, resources, and guidelines. *Computers and Education: Artificial Intelligence*, *7*, 100326. https://doi.org/10.1016/j.caeai.2024.100326

Wood, D. & Moss, S. H. (2024). Evaluating the impact of students' generative ai use in educational contexts. *Journal of Research in Innovative Teaching & Learning*, *17*(2), 152-167. https://doi.org/10.1108/jrit-06-2024-0151

Yusuf, A., Bello, S., Pervin, N., & Tukur, A. K. (2024). Implementing a proposed framework for enhancing critical thinking skills in synthesizing ai-generated texts. *Thinking Skills and Creativity*, *53*, 101619. https://doi.org/10.1016/j.tsc.2024.101619

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), 1-27. https://doi.org/10.1186/s41239-019-0171-0

Zimmerman, B. J., & Schunk, D. H. (Eds.). (2011). *Handbook of self-regulation of learning and performance*. Routledge.