Consuegra-Fernández, M., Sanz-Aznar, J., Burguera-Serra, J. G., and Caballero, J. J. (2024). ChatGPT: The Dilemma of the Authorship of Graded Assignments in Higher-Education. *Revista de Investigación Educativa, 42*(2).
DOI: https://doi.org/10.6018/rie.565391

ChatGPT: The Dilemma of the Authorship of Graded Assignments in Higher-Education

ChatGPT: el dilema sobre la autoría de las actividades evaluables en educación universitaria

Marta Consuegra-Fernández*[1] , Javier Sanz-Aznar*, Joan Gabriel Burguera-Serra* and Juan José Caballero Molina*.

*Faculty of Philology and Communication. Universitat de Barcelona (Spain)

Abstract

The emergence of ChatGPT poses new challenges in the educational field. Among them, the open discussion on the potentially negative consequences that the program's use may generate in the learning and evaluation processes of students. The present study investigates the level of knowledge and perception of ChatGPT among university educators, as well as their proficiency in discerning student-authored texts from those generated by artificial intelligence. For this purpose, 51 professors at the University of Barcelona, specializing in communication and philology, were presented with a sample of texts extracted from an authentic academic assignment that included versions written by students themselves, together with outputs generated ad hoc by ChatGPT. The accuracy rate of the authorship assignment performed by teachers was 31%, a value that reveals a new obstacle in teaching, learning, and evaluation processes in higher education. Additionally, there was a tendency for ChatGPT-generated texts to be rated more favorably than those written by the students themselves. Finally, the article presents several

[1] **Correspondence:** Marta Consuegra-Fernández, mconsuegra@ub.edu, Universitat de Barcelona Gran Via de les Corts Catalanes, 585, 08007 Barcelona.

suggestions aimed at anticipating the potential impact of the unethical use of artificial intelligence on the development of skills and abilities among university students.

Keywords: ChatGPT; Natural Language Processing; Written Communication; University Education.

Resumen

La irrupción de ChatGPT plantea nuevos desafíos en el ámbito educativo. Entre ellos, destaca el debate abierto en torno a las consecuencias -potencialmente negativas- que los usos de la herramienta pueden generar en los procesos de aprendizaje y evaluación del alumnado d. El siguiente trabajo explora tanto el grado de conocimiento y percepción sobre ChatGPT del profesorado universitario, como la capacidad para distinguir trabajos de autoría humana de otros originados por la inteligencia artificial. Para ello, 51 docentes de la Universidad de Barcelona, de áreas de conocimiento asociadas a la comunicación y la filología, tuvieron acceso a los textos resultantes de una actividad académica real, a partir de versiones redactadas por los propios alumnos, y de otras generadas ad hoc a través de ChatGPT. Los resultados revelaron un porcentaje de acierto en la asignación de autoría del 31 %, un valor que evidencia un nuevo obstáculo en los procesos de enseñanza, aprendizaje y evaluación en la educación superior. Paralelamente, se observó que tendían a otorgar una valoración más positiva a las muestras elaboradas por ChatGPT frente a aquellas redactadas por el alumnado. Finalmente, el artículo recoge una serie de propuestas para anticipar el impacto que podría tener un uso deshonesto de la inteligencia artificial en la adquisición de competencias y habilidades del alumnado universitario.

Palabras clave: ChatGPT; Procesamiento del lenguaje natural; comunicación escrita; educación universitaria.

Introduction and objectives

Advances in the fields of artificial intelligence and machine learning (Chiche and Yitagesu, 2022; Bharadiya, 2023) have contributed significantly to the growth of computational linguistics. One of the most advanced techniques in the last two decades is Natural Language Processing (NLP) which, through computational models and algorithms, allows the analysis of free texts and the extraction of relevant information (Locke et al., 2023). (Locke et al., 2021).. This utility deepens both the ability to understand and to generate natural language itself, so that its applications provide an interface for users to formulate questions or instructions, and access information in the form of computer programmes designed to simulate, among others, a human conversation and perform a wide variety of tasks, the so-called *chatbots* (Luo et al., 2022)..

PLN has shown great potential with the launch of ChatGPT, an OpenAI machine learning model that uses algorithms based on more than 150 billion features of human language[2] and responds to user *prompts*. ChatGPT uses a massive learning model based on texts accessible on the internet, whose automatic review allows it to deduce and

[2]Available at https://openai.com/blog/chatgpt/

recognise shared linguistic structures. This learning is updated as the programme accesses and compares new texts with previously apprehended patterns, thus perfecting its PLN ability. There are currently several versions of the programme, including GPT-3.5, released in November 2022, which is a free prototype test version, and GPT-4, released in March 2023, which is only available on a paid subscription basis[3] .

Although PLN has been developing since the end of the 20th century (Manning and Schutze, 1999)(Manning and Schutze, 1999), GPT technology (*Generative Pre-training Transformer*) (Zhu and Luo, 2022) is a revolutionary qualitative breakthrough and has been considered by some authors as a disruptive innovation (Dowling and Lucey, 2022). (Dowling and Lucey, 2023).. Its use has become so popular that it has become a mass resource. Proof of this is the increasing media coverage it has received since its launch, to the point of occupying a prominent role in the public agenda (Fig. 1).

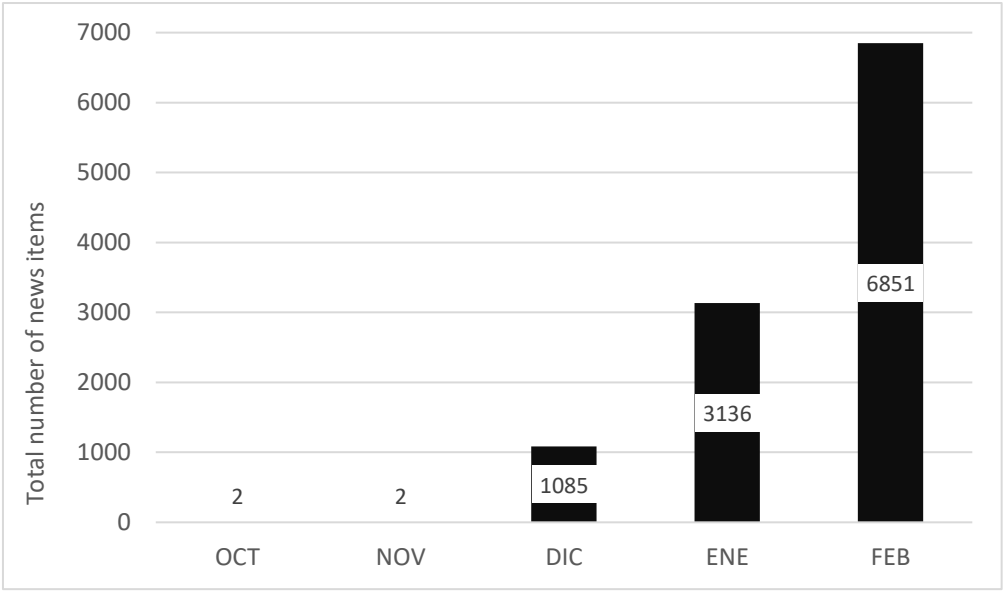

Figure 1. News obtained with the keyword search "ChatGPT" in the Spanish contemporary press archive (Source: *MyNews*).

This media notoriety has attracted the interest of users from many different areas (Rudolph et al., 2023). (Rudolph et al., 2023) and especially from the student community (Haque et al., 2022). (Haque et al., 2022)..

In the education sector, ChatGPT is used by users for a variety of tasks including text composition or information gathering and processing. However, unlike in sectors such

[3]Available at https://openai.com/blog/chatgpt/

as digital marketing, where the use of ChatGPT is more widely accepted, its use for educational purposes is controversial (Baidoo-Anu and Owusu, 2023).. While some perceive PLN systems as a possible didactic tool that is efficient, accessible and able to reduce costs for educational institutions (Heller et al., 2005), the use of PLN systems in education is controversial (Baidoo-Anu and Owusu, 2023). (Heller et al., 2005; Perez et al., 2020; Tallyn et al., 2018; Villegas-Ch et al., 2020)Others consider that it cannot replace human interaction, which is fundamental for the social and communicative development of learners (Butnaru et al., 2018). (Butnaru et al., 2021; Sharp and Theiler, 2018).. On the other hand, ChatGPT has also been criticised for the potential presence of ideological biases in the algorithm, which, as has happened on other occasions, would pose a threat to the well-being of some groups (Doshi et al., 2021). (Doshi et al., 2023).. However, one of the most controversial issues regarding ChatGPT lies in the possible misappropriation, in terms of authorship, that students may make of the texts and content produced by the tool. If this were the case, it would not only call into question the assessment process itself on the part of the teachers (Baidoo-Anu and Owusu, 2023)but it would also seriously compromise the learning and critical thinking skills that students must demonstrate (García-Peñalvo, 2023). (García-Peñalvo, 2023).. At this point it is also worth noting the growing trend of non-face-to-face academic assignments and assessments (Gómez and Alende, 2023). (Gómez and Alende, 2022; Sáiz-Manzanares et al., 2022).which inevitably entail less control of authorship by the teacher, as well as a greater risk of plagiarism and academic fraud. Proof of this is the popularity of plagiarism detection tools, which are becoming more and more widespread in university pre-assessment processes (Khalil and Er, 2023)..

This fear is based on evidence gathered by users who have submitted ChatGPT to various assessment tests. Among others, it seems that the programme is able to pass law exams (Bommarito and Katz, 2022; Choi et al., 2023). (Bommarito and Katz, 2022; Choi et al., 2023) or obtain a medical licence to practice in the United States (Kung et al., 2023). (Kung et al., 2023).. In Spain, it has also been able to pass the cut-off mark of the MIR exam in 2022 (Carrasco et al., 2022). (Carrasco et al., 2023).. However, at the time of writing this manuscript, only one *pre-print* article has been published that has explored, with a scientific methodology, the real risks of a possible illegitimate use in the scientific-academic environment (Else, 2023). (Else, 2023). A group from the Feinberg School of Medicine of Northwestern University in the United States (Gao et al. (Gao et al., 2022) has recently shown that ChatGPT is able to write plausible scientific summaries from invented content. Specifically, after a careful reading, a group of researchers, previously warned of the objective of the study, were able to correctly assign the artificial authorship of the texts in 68% of the cases. (Gao et al., 2022)..

This scenario determines the need to assess its possible impact, as well as the relevance of articulating preventive measures if necessary. While it is true that computational systems have been appearing that can detect whether a text in English has been generated using PLN (such as GPTZero or Contentatscale AI Content Detector), there is no evidence of the effectiveness of these tools in Spanish.

In this context, the hypothesis of this research argues that university professors do not have effective mechanisms to distinguish texts produced by students from those

generated by artificial intelligence. If this hypothesis is validated, the impact in terms of risk derived from the irruption of ChatGPT in higher education would call into question the evaluation processes of certain types of academic activities.

The objectives of this research are:

- To examine the general qualitative assessment that university teachers give to anonymised texts produced either by students or by ChatGPT.
- To evaluate the level of accuracy of teachers in assigning authorship (students vs. ChatGPT) of these texts.
- To analyse the degree of knowledge, perception of the impact and needs of teaching staff with regard to the use of artificial intelligence in a university context.

Method

The research has been carried out in collaboration with a sample of active teachers from different degrees of the Faculty of Philology and Communication (FiC) of the University of Barcelona (UB) and responds to the need to provide empirical evidence to some of the reflections that are being held in relation to the incursion of ChatGPT in the education sector. This is a cross-sectional, descriptive and analytical design study that combines quantitative and qualitative methodology, and uses the questionnaire technique with items adapted from previous references. Briefly, teachers with specific training and pedagogical skills in language, literary analysis and communication were asked, after reading three anonymised texts corresponding to a real evaluative activity, to rank them qualitatively and, subsequently, to indicate whether the texts under analysis had been produced by students or by ChatGPT.

Population and Sample

For the present research, a sample of 51 lecturers from the UB's Faculty of Communication and Culture who were part of the Teaching and Research Staff during the academic year 2022-23 was obtained. The lecturers are attached to the Departments of Classical, Romance and Semitic Philology; Hispanic Philology, Literary Theory and Communication; Catalan Philology and General Linguistics; and Modern Languages and Literatures and English Studies. Their academic and professional profile accredits a broad knowledge of language, communication and culture, and a great command of the analysis and interpretation of different types of texts. Their training and dedication to both research and teaching make them specialists in text analysis.

The sample (Table 1) is made up of teachers from the fields of Communication, Philology, Literary Studies and/or Linguistics, aged between 32 and 70, and with an average of 21 years of heterogeneous teaching experience.

Table 1

Description of the sample of teachers (N=51)

Variables	Descriptive values
Age	*(years)*
1. Average	50,39
2. Standard deviation	±9,37
3. Fashion	47
4. Age range	[32,71]
Gender	*(individuals)*
1. Female	27 (59,94%)
2. Male	22 (43,14%)
3. Not determined	2 (3,92%)
Area of knowledge	*(individuals)*
1. Philology, literary studies and/or linguistics	42 (82,35%)
2. Audiovisual communication and advertising	4 (7,84%)
3. Journalism	5 (9,80%)
Years dedicated to university teaching	*(years)*
1. Average	21,06
2. Deviation	±10,28
3. Fashion	15
4. Range of experience	[4,49]

Instrument

In order to carry out the experiment, the literary review was chosen as the textual genre to be analysed by the teaching team, and a total of 6 reviews were obtained, 3 of them written by ChatGPT and 3 by students.

As a language model trained through thousands of public texts on the internet, ChatGPT is capable of answering multiple questions of different kinds and also of producing academic texts of different typologies[4] . However, in order to avoid the influence of variations inherent in the pattern and style of the exercise, only literature reviews were included. The review fulfils the specific requirements of an evaluable delivery present in the syllabus of all FiC degrees, so that the teachers participating in

[4]Available at https://openai.com/blog/chatgpt/

the research are familiar with its reading and evaluation. It is also a complex task that seeks to ascertain the student's ability to carry out a critical and objective analysis of a cultural product.

The student-written reviews included in the experiment come from real samples developed in the 2021-2022 academic year, within the framework of the subject "Genres and formats of written communication" of the FiC Degree in Communication and Cultural Industries. This is a compulsory subject in the second year. Samples were chosen from the 21-22 academic year, as they predate the existence of ChatGPT, in order to avoid the possibility that the reviews might already have been fully or partially developed using ChatGPT. In addition, a specific submission of a single subject was defined to ensure that each text came from a different student, thus guaranteeing the heterogeneity of the sample.

After anonymising the collected texts, 40 in total, the selection was made according to the following inclusion criteria: (*i*) texts each dealing with a different work, and (*ii*) texts receiving a minimum rating of "notable". The selection also ruled out (*i*) texts longer than 600 words, (*ii*) texts with their own subtitles, images, graphics or other structural or design elements, (*iii*) texts with a structural division of less than 3 paragraphs, (*iv*) texts written in the first person singular, and (*v*) texts with a plagiarism rate of more than 1 % according to the Urkund system. Of the 7 resulting reviews, 3 were randomly selected (available in the Appendix).

The 3 texts produced by ChatGPT in its GPT-3 version were obtained through the OpenAI platform in response to the following instruction: "*Write a 500-word literary review in Spanish about the book...*" (see Appendix). In all three cases, the title of the work to be reviewed was specified, which, in order to be consistent with the books chosen by the students, also corresponded to works of journalistic narrative (Table 2). The instruction given to ChatGPT is identical to the statement of the evaluation activity carried out by the students in order to establish the same basic conditions of execution. While it is true that the platform allows for the inclusion of further instructions that add complexity or quality to the resulting text, the choice of a basic instruction avoids the influence of experience and user knowledge operating ChatGPT, features that could act as confounding variables.

Table 2

Literary works narrated in the three reviews included in the research and the number of teachers who reviewed each of them.

Work	Author	Year of issue	Author of the review	No. of teachers exposed to each sample
The chauffeur's son	Jordi Amat i Fusté	2020	ChatGPT	25

In cold blood	Truman Capote	1965	ChatGPT	23
Inshallah	Orianda Fallaci	1992	ChatGPT	28
She said	Jodi Kantor and Megan Twohey	2022	Student	25
The journalist and the killer	Janet Malcom	2004	Student	28
Ebony	Ryszard Kapuściński	2006	Student	24

In order to explore teachers' perception of the reviews, a questionnaire (available in the Appendix) was developed using Pavlovia[5] , an open access platform that offers tools for creating, running and collecting data online. The questions were structured as follows: (*i*) reading and ranking of the reviews on the basis of a general qualitative assessment in a broad sense (quality of content, formal aspects, etc.), (ii) attribution of the reviews to the authors, (iii) the quality of the reviews, (iv) the quality of the content, and (v) the quality of the reviews.), (*ii*) attribution of the authorship of each text to a student or to artificial intelligence (the impossibility of defining authorship was considered as a response), (*iii*) socio-demographic and educational characterisation of each participant, and (*iv*) knowledge and self-perception, assessed using a discrete Likert scale of 6 grades from 1 (strongly disagree) to 6 (strongly agree), regarding teachers' knowledge of the ChatGPT tool, their opinion of its usefulness in the education sector, distrust and perception of threat, the need for specific training on it, and their anticipation of its future scope. In order to simulate a real evaluation context and to avoid bias or pre-conditioning teachers in the assignment of authorship, no reference was made to the purpose of the research. Moreover, teachers only had access to the questions regarding their knowledge and perception of ChatGPT once the qualitative evaluation and attribution of authorship of the texts had been completed, without the possibility of backtracking to allow them to repeat the process and correct their initial assessment.

The 9 items assessed on the Likert scale were adapted from the publication by Escoda and Conde (2016) and, to ensure their reliability, the Cronbach's alpha values corresponding to each item were calculated. In all cases, values equal to or higher than 0.7 were obtained, which is the reliability threshold established for small samples (Bujang et al. (Bujang et al., 2018)..

Prior to its distribution, the tool was checked for correct functioning by means of a pilot test carried out on 4 subjects who met the same requirements as the sample participants and who were not part of the study.

[5]Available at https://pavlovia.org/

Data collection and analysis procedure

Each teacher analysed three reviews which were assigned through a pseudo-random process to avoid combinatorial bias, ensuring that (*i*) each participant read different texts, (*ii*) at least one or two of these texts were written by ChatGPT, and one or two by students, (*iii)* that all texts were read and reviewed a similar number of times, (*iv*) that all texts were shown the same number of times in first, second or third place, and finally, (*v*) that all situations in which two texts by one author were presented were the same as those in which two texts by the other author were presented.

In all cases, subjects were informed that their participation was anonymous and voluntary, and that all data collected would be used only by the UB, excluding the transfer to third parties. They were also informed of the commitment to comply with the provisions of Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April on the protection of individuals with regard to the processing of their personal data and the free movement of such data and Organic Law 3/2018 of 5 December on the protection of personal data and guarantee of digital rights.

Participants were selected by means of non-probabilistic convenience sampling. The questionnaire was sent on 25 February 2023 via a UB email account belonging to the project to 163 lecturers at the UB's Faculty of Science and Technology, out of a total of 358[6] , who were part of the Teaching and Research Staff during the academic year 2022-23. A participation rate of 31.29% was obtained, equivalent to 51 completed questionnaires that constituted the final representative sample (with a confidence level of 95% and a margin of error of 13%).

Responses were collected and analysed using IBM SPSS *Statistics* v.26. The numerical results of the assignment and assessment of the texts are presented in absolute numbers and in percentages of relative frequencies. To analyse the associations between the two nominal variables "ChatGPT authorship" and "student authorship", Pearson's Chi-square components and Cramer's indices were calculated from the observed and expected frequencies to quantify the degree of relationship between 0 and 1. Statistical significance was set at $p<0.05$ and the confidence level at 95 %.

The Likert scale responses are represented by the mean (M), median (Md) and standard error of the mean (SEM) of the total number of responses. In addition, items of a qualitative nature where a written response was requested were also analysed.

Results and discussion

The reviews were reviewed a total of 153 times by teachers (Figure 2). Following these readings, in 48 cases (31.37 %) the authorship was assigned to students and on 43 occasions to ChatGPT (28.10 %), while in as many as 62 cases teachers claimed "not being able to identify authorship" (40.52 %).

[6]Retrieved 21 February 2023 from the website of the Facultat de Filologia i Comunicació de la Universitat de Barcelona https://www.ub.edu/portal/web/filologia-comunicacio

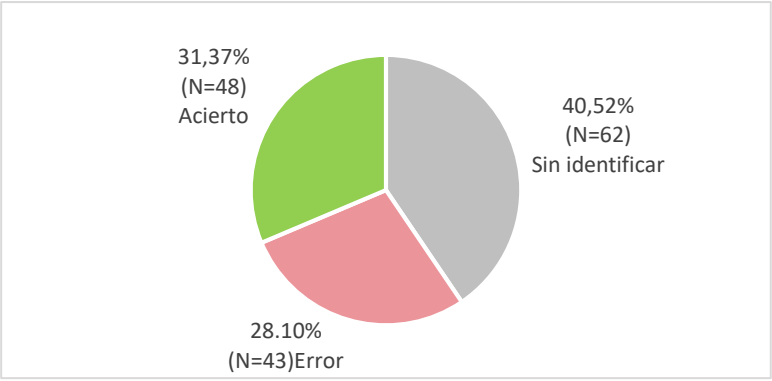

Figure 2. Percentage and absolute numbers of correct (hit), incorrect (miss) and unidentified assignments in relation to the authorship of the texts.

As shown in Figure 2, the hit rate for these assignments was 31.37%, which is lower than the 68% hit rate obtained by the research team analysing scientific abstracts in the study by Gao et al. (2022).. This first finding, together with the high number of unidentified abstracts, reveals the real risk of potential fraud going undetected. The difference with respect to the study by Gao et al. (2022) is because in their case, the research team informed the reviewers in advance of the existence of ChatGPT-generated summaries, whereas in the present study, the faculty were unaware of the purpose of reading the texts. They were also not allowed to go back through the questionnaire to carry out a more detailed reading and a specific search for signs of writing with PLN systems. This methodological difference between the two studies is fundamental to understanding the percentage of authorship identification of the texts reviewed by the teachers in our research, and highlights the need for teachers to be aware of the programme and the consequences of its use.

Table 3

Analysis of hits, misses and unidentified authorship assigned to students and ChatGPT

	% Hit (N)	**Error (N)**	**Y/Y/iden (N)**	**Pearson Chi-square**	**V for Cramér**	**Sig. bilater.**
ChatGPT	13,1% (20)	13,7% (21)	22,9% (35)	2,384	0,125	0,304
Students	18,3% (28)	14,4% (22)	17,6% (27)			

In addition, in-depth analysis of the data revealed that hits, misses and unidentified assignments were evenly distributed across both student and ChatGPT reviews (Table 3), suggesting that there were no distinguishing features of writing by one author or the other. However, when asked about the reasons why they had linked certain reviews to artificial intelligence, 60.8% responded with reasons that have been grouped by coinciding themes in Table 4, while the rest did not answer or did not know.

Table 4

Reasons for assigning texts to ChatGPT grouped by subject matter

Themes	**Responses from participants**
Structure	"The structure is too mechanised. Although in all three texts there is a similar (problematic) structure".
Written quality	"... compared to the other two, shows poorer written expression".
Reiterations	"... is essentially reiterative and each paragraph is syntactically constructed in a very similar way".
Punctuation errors	"...the punctuation and the complexity of the sentences give me this doubt." "...some mismatches and commas."
Enunciative modality	"The neutral and impersonal character of the style", "The expression of one of the texts seemed to me merely informative, constructed with typified stylistic resources, without much personality".
Literal translations	"The sequence 'the Catalan family' is used (...) with initial capital letters, which is suspicious that it comes from an original in another language, possibly English. ", "...lexical uses more typical of translated works".

Among the arguments that the teaching group claimed as discriminatory elements between the students' texts and those of the artificial intelligence, the presence of literal reiterations and paraphrased equivalent structures was found to be the most frequent aspect (20 %). This is one of the most notable features of ChatGPT's own writing style that can even evade plagiarism detection systems such as GPTZero, GPT-2 Output Detector or AI Detector (Anderson et al., 2023). Less frequently, the teaching group argued the presence of neutral and not very expressive enunciative modalisation, errors

of discursive coherence (punctuation, textual sequencing, etc.) and an apparent lexical availability inconsistent with the presumable lexical availability of the students. Moreover, the mechanised text structure, one of the distinctive elements of ChatGPT writing, already observed in previous publications (Gao et al., 20. (Gao et al., 2022)was mentioned twice. In general, it was observed that the teachers' criteria coincide with each other as well as with those reported in the limited literature available to date (Gao et al., 2022). (Gao et al., 2022).. However, they were not always conclusive enough to correctly determine authorship and sometimes even led to misidentification.

After observing that there is a real risk of misidentification of the authorship of the reviews, we proceeded to analyse the qualitative prioritisation of the texts by the teachers.

Table 5

Ranking of the reviews on the basis of the overall qualitative rating

	1ST % (N)	**% 2º (N)**	**3º (N)**	**Pearson Chi-square**	**V for Cramér**	**Sig. bilater.**
ChatGPT	18,3% (28)	19,0% (29)	12,4% (19)			
				4,758	0,176	0,09
Students	15,0% (28)	14,4% (22)	20,9% (32)			

The results (Table 5) showed a tendency to prioritise texts written by ChatGPT over texts written by students. However, the study sample may not be sufficient to detect a small effect (V Cramer=0.176), as indicated by the resulting significance value (p=0.09). Thus, , although future research with a larger sample may be necessary to analyse the significance of the bias , and even to include numerical assessments, the observed trend cannot be ignored because of the potential negative impact it may have. Overall, it is concluded that not only is there a risk in the assignment of authorship, but that texts generated by ChatGPT could be rated higher than those produced by students. If this is the case, the fraudulent use of the tool would represent an evaluative advantage and could promote such practices among students (Agud, 2014; Díez-Martínez, 2015)..

It is not the first time that technological innovation has had an impact on the field of education and has challenged traditional teaching and learning methodologies. However, current curricula and teaching plans integrate information and communication technology (ICT) competences (Domingo-Coscollola et al., 2020).. This concerns not only the digital literacy of learners, but also the so-called Digital Competence in Teaching (DTT) (Hall et al., 2014) which encompasses the teaching skills and knowledge to use these methodologies and tools as teaching resources in the classroom. In this sense, it is important to know the opinion and knowledge of teachers about PLN systems such as ChatGPT that could be part of this technological

transformation. Moreover, knowledge of the tool alerts teachers to its possible illicit applications and anticipates them to recognise fraudulent authorship. (Gao et al., 2022)..

In relation to this aspect, the analysis of the self-perception questionnaire answered by the teaching staff revealed that they currently have little practical knowledge of the tool (Table 6, *Item 2*), probably due to its recent appearance, and that they express a clear desire to receive specific training in this respect (Table 6, *Item 7*). In line with these results, other studies point to the fact that university educational practice is not in tune with the current digital reality (Domingo-Coscollola et al., 2020; Sancho-Gil et al., 2017) and stress the need to enrich the CDD training of teaching staff so that they can adequately develop their teaching tasks (Cervera et al., 2016). (Cervera et al., 2016; Cuartero et al., 2016)..

On the other hand, the participants also do not have a clear opinion about ChatGPT, nor about whether it represents a threat to university teaching. The trend observed is that the teaching staff consider that the tool will have a greater potential in the future (Table 6, *Item 8*), despite the fact that the most recent literature shows that the current predominant use belongs to the student community (Haque et al., 2022). (Haque et al., 2022)..

Table 6

Knowledge and perception assessed by Likert questions from 1 (strongly disagree) to 6 (strongly agree) to the participating teachers (N=51).

***Item* for each question**	**Mean (M)**	**Median (Md)**	**Mean error (SEM)**
Item 1. You are receptive to the use of programmes such as ChatGPT in your academic activity.	2,82	3	1,57
Item 2. Have you used ChatGPT	1,69	1	1,42
Item 3. Do you think these programmes pose a threat to university teaching?	4,04	4	1,60
Item 4. Do you think it is urgent to change the assessment in your subjects?	3,92	5	1,94
Item 5. Do you think action should be taken on this issue?	4,31	5	1,66
Item 6. Do you think it can be a useful tool for university education?	3,37	4	1,56
Item 7. Do you think that university teaching staff should receive specific training on these new tools?	4,76	6	1,69

Item 8. Do you think students will use this type of tool professionally?	4,90	5	1,33
Item 9. Do you think your students use these types of resources in their academic work?	3,33	3	1,60

It is worth noting that no significant differences associated with age, gender, education and years of teaching have been identified (analysis using *Student's t-test* for parametric variables and *Mann-Witney* for non-parametric variables, in each case). Apparently, any categorisation with the aforementioned variables behaves in the same way as the overall results observed in terms of identification of authorship, qualitative prioritisation of texts, and knowledge and self-perception measured through the Likert scale. The present research has focused on the study of a very homogeneous sample of teachers with specific training and pedagogical skills in language, communication and literary analysis. However, we do not rule out the existence of possible differences between age groups, gender, training and years of teaching of the participants, which could be studied if the sample were enlarged and diversified, including teachers from other universities and, above all, from other areas of knowledge.

Conclusions

This paper presents one of the first assessments of the risk associated with illegitimate use of ChatGPT technology and questions the assessment process of certain assignments, while suggesting the existence of interference in student learning. The research carried out corroborates the research hypothesis: the evidence obtained shows that university lecturers in the subject areas in question do not have mechanisms to distinguish between texts generated by ChatGPT and texts written by students. Moreover, in general terms, teachers are still unfamiliar with this type of tool and are unaware of its potential practical use.

However, the irruption of PLN systems, such as ChatGPT, in the educational sector is an inevitable reality that will presumably consolidate over time, as has happened with other technological innovations that were also perceived as *disruptive* at the time. In this sense, it does not seem feasible or even positive to categorically prohibit the use of artificial intelligence for pedagogical purposes and, instead, the mechanisms for integrating it into the educational framework should be explored, as a new resource, in order to evaluate its benefits in the same way that its potential limitations and damages are also anticipated. It also seems appropriate to propose some actions, namely (*i*) informing teachers and students about the existence of the different tools that use PLN and applications derived from artificial intelligence so that they are aware of both their potential uses and drawbacks, (*ii*) training teachers and students to use these resources correctly and ethically, and to avoid deviations that interfere with the development of learning, (*iii*) integrating new technologies and tools into teaching plans, (iv) providing guides and resources to standardise the use of PLN, (v) providing guides and resources to standardise the use of PLN and applications derived from artificial intelligence, (vi)

providing guides and resources to standardise the use of PLN and applications derived from artificial intelligence, (vii) providing guides and resources to standardise the use of PLN and applications derived from artificial intelligence, (*iv)* limit their use in cases where such tools severely condition the acquisition of skills or competences that are the object of learning-assessment, which could require attendance and manual writing in certain cases, and (*v*) rethink assessment tasks or activities that may have become obsolete in the current scenario of technological transformation, and replace them with others that include the new digital challenges.

The methodology has some limitations that need to be considered. Firstly, the study includes only texts produced by the ChatGPT programme (version GPT-3), so the implications may not be representative of its updates or of other PLN systems. Secondly, the results are obtained from the specific study around the genre 'literary review' and might differ from other discursive genres. Thirdly, the quality of the texts generated by ChatGPT depends to a large extent on the clarity and concreteness of the information provided in the user instruction. In this sense, more complex instructions could have resulted in different reviews which, foreseeably, could have altered the opinion of the teachers.

A possible future line of analysis could replicate similar research after providing specific training to teaching teams on the operation of the ChatGPT tool and its PLN system in order to test the expected increase in the identification of authorship of the texts analysed.

Acknowledgements

This work has been carried out within the framework of the research promoted by the Grup d'Innovació en Comunicació i Emprenedoria of the University of Barcelona. We would like to thank the collaboration of the teaching staff of the FiC of the UB for their participation in the study and the interest shown in the research.

References

Agud, J. L. (2014). Fraud and plagiarism in the career and profession. *Revista Clínica Española*, *214*(7), 410-414. https://doi.org/10.1016/J.RCE.2014.03.007

Anderson, N., Belavy, D. L., Perle, S. M., Hendricks, S., Hespanhol, L., Verhagen, E., and Memon, A. R. (2023). AI did not write this manuscript, or did it? Can we trick the AI text detector into generated texts? The potential future of ChatGPT and AI in Sports & Exercise Medicine manuscript generation. *BMJ Open Sp Ex Med*, *9*, 1568. https://doi.org/10.1136/bmjsem-2023-001568

Baidoo-Anu, D., and Owusu, L. (2023). Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning. *SSRN.* https://doi.org/10.2139/SSRN.4337484

Bharadiya, J. (2023). A Comprehensive Survey of Deep Learning Techniques *Natural Language Processing. European Journal of Technology, 7*(1), 58-66. https://doi.org/10.47672/ejt.1473

Bommarito, M. J., and Katz, D. M. (2022). GPT takes the Bar Exam. *SSRN*. http://dx.doi.org/10.2139/ssrn.4314839

Bujang, M. A., Omar, E. D., & Baharum, N. A. (2018). A Review on Sample Size Determination for Cronbach's Alpha Test: A Simple Guide for Researchers. *The Malaysian Journal of Medical Sciences : MJMS*, *25*(6), 85. https://doi.org/10.21315/MJMS2018.25.6.9

Butnaru, G. I., Haller, A. P., Dragolea, L. L., Anichiti, A., and Hârșan, G. D. T. (2021). Students' Wellbeing during Transition from Onsite to Online Education: Are There Risks Arising from Social Isolation? *International Journal of Environmental Research and Public Health*, *18*(18). https://doi.org/10.3390/IJERPH18189665

Carrasco, J. P., García, E., Sánchez, D. A., Porter, E., De La Puente, L., Navarro, J., and Cerame, A. (2023). Is "ChatGPT" capable of passing the 2022 MIR exam? Implications of artificial intelligence in medical education in Spain. *Revista Española de Educación Médica*, *4*(1). https://doi.org/10.6018/edumed.556511

Cervera, M. G., Martínez, J. G., and Mon, F. M. E. (2016). Digital competence and digital competence in teaching: an overview of the state of the art. *RiiTE Revista Interuniversitaria de Investigación En Tecnología Educativa*, 2529-9638. https://doi.org/10.6018/RIITE2016/257631

Chiche, A., and Yitagesu, B. (2022). Part of speech tagging: a systematic review of deep learning and machine learning approaches. *Journal of Big Data, 9*(1), 1-25. https://doi.org/10.1186/s40537-022-00561-y

Choi, J. H., Hickman, K. E., Monahan, A. B., and Schwarcz, D. (2023). ChatGPT goes to Law School. *SSRN.* https://dx.doi.org/10.2139/ssrn.4335905

Cuartero, M. D., Gutiérrez, I., Paz, M., Espinosa, P., and Clave, P. (2016). Análisis Conceptual de Modelos de Competencia Digital del Profesorado Universitario / Conceptual analysis of digital competence models of university teacher. *Revista Latinoamericana de Tecnología Educativa - RELATEC*, *15*(1), 97-114. https://doi.org/10.17398/1695-288X.15.1.97

Díez-Martínez, E. (2015). Academic dishonesty of students and teachers. Their contribution to moral disengagement and social corruption. *Sinéctica*, *44*, 14. https://dialnet.unirioja.es/servlet/articulo?codigo=8239864&info=resumen&idioma=ENG

Domingo-Coscollola, M., Bosco, A., Segovia, S. C., and Valero, J. A. S. (2020). Fostering teaching digital competence at university: Perception of students and teachers. *Revista de Investigación Educativa*, *38*(1), 167-182. https://doi.org/10.6018/RIE.340551

Doshi, R. H., Bajaj, S. S., and Krumholz, H. M. (2023). ChatGPT: Temptations of Progress. *The American Journal of Bioethics.* https://doi.org/10.1080/15265161.2023.2180110

Dowling, M., and Lucey, B. (2023). ChatGPT for (Finance) research: The Bananarama Conjecture. *Finance Research Letters*, *53*, 103662. https://doi.org/10.1016/J.FRL.2023.103662

Else, H. (2023). Abstracts written by ChatGPT fool scientists. *Nature*, *613*(7944), 423. https://doi.org/10.1038/D41586-023-00056-7

Escoda, A. P., and Conde, M. J. R. (2016). Evaluation of the self-perceived digital competences of Primary Education teachers in Castilla y León (Spain). *Revista de Investigación Educativa*, *34*(2), 399-415. https://doi.org/10.6018/RIE.34.2.215121

Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., and Pearson, A. T. (2022). Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers. *BioRxiv*, 2022.12.23.521610. https://doi.org/10.1101/2022.12.23.521610

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 (EKS)*, *24*, e31279. https://doi.org/10.14201/EKS.31279

Gómez, C., and Alende, S. (2022). Rutinas y tareas del profesorado universitario durante la covid-19, Estudio de caso de la facultad de ciencias sociales y comunicación de la Universidad de Vigo. *Research: Culture, Science and Technology*, *27*, 36-48. https://dialnet.unirioja.es/servlet/articulo?codigo=8626674&info=resumen&idioma=SPA

Hall, R., Atkins, L., & Fraser, J. (2014). Defining a self-evaluation digital literacy framework for secondary educators: the DigiLit Leicester project. *Research in Learning Technology*, *22. https://doi.*org/10.3402/RLT.V22.21440

Haque, M., Dharmadasa, I., Tasnim Sworna, Z., Namal Rajapakse, R., and Ahmad, H. (2022). "I think this is the most disruptive technology" Exploring Sentiments of ChatGPT Early Adopters using Twitter Data. *Arxiv.* https://doi.org/10.48550/arXiv.2212.05856

Heller, B., Heller, B., Proctor, M., Mah, D., Jewell, L., and Cheung, B. (2005). Freudbot: An Investigation of Chatbot Technology in Distance Education. *EdMedia + Innovate Learning*, *2005*(1), 3913-3918. https://www.learntechlib.org/primary/p/20691/

Khalil, M., and Er, E. (2023). Will ChatGPT get you caught? Rethinking of Plagiarism Detection. *Arxiv.* https://doi.org/10.48550/arXiv.2302.04335

Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., and Tseng, V. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digital Health*, *2*(2), e0000198. https://doi.org/10.1371/JOURNAL.PDIG.0000198

Organic Law 3/2018, of 5 December, on the Protection of Personal Data and Guarantee of Digital Rights. *Official State Gazette, 294,* 6 December 2018. https://www.boe.es/buscar/act.php?id=BOE-A-2018-16673

Locke, S., Bashall, A., Al-Adely, S., Moore, J., Wilson, A., & Kitchen, G. B. (2021). Natural language processing in medicine: A review. *Trends in Anaesthesia and Critical Care*, *38*, 4-9. https://doi.org/10.1016/J.TACC.2021.02.007

Luo, B., Lau, R. Y. K., Li, C., and Si, Y. W. (2022). A critical review of state-of-the-art chatbot designs and applications. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *12*(1), 1-26. https://doi.org/10.1002/WIDM.1434

Manning, C., and Schutze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press. http://nlp.stanford.edu/fsnlp/

Pérez, J. Q., Daradoumis, T., and Puig, J. M. M. (2020). Rediscovering the use of chatbots in education: A systematic literature review. *Computer Applications in Engineering Education*, *28*(6), 1549-1565. https://doi.org/10.1002/CAE.22326

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April on the protection of individuals with regard to the processing of their personal data and on the free movement of such data and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union, L119*, 4 May 2016. https://eur-lex.europa.eu/legal-content/es/ALL/?uri=CELEX%3A32016R0679

Rudolph, J., Tan, S., and 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

Sáiz-Manzanares, M.C., Casanova, J.R., Lencastre, J.A., Almeida, L., and Martín-Antón, L.J. (2022). Student satisfaction with online teaching in times of COVID-19. *Comunicar*, *30*(70). https://doi.org/10.3916/C70-2022-03

Sancho-Gil, J. M., Sánchez-Valero, J. A., & Domingo-Coscollola, M. (2017). Research-based insights on initial teacher education in Spain. *European Journal of Teacher Education*, *40*(3), 310-325. https://doi.org/10.1080/02619768.2017.1320388

Sharp, J., & Theiler, S. (2018). A Review of Psychological Distress Among University Students: Pervasiveness, Implications and Potential Points of Intervention. *International Journal for the Advancement of Counselling*, *40*(3), 193-212. https://doi.org/10.1007/s10447-018-9321-7

Tallyn, E., Fried, H., Gianni, R., Isard, A., and Speed, C. (2018). The ethnobot: Gathering ethnographies in the age of IoT. *Conference on Human Factors in Computing Systems - Proceedings*, 604, 1-13. https://doi.org/10.1145/3173574.3174178

Villegas-Ch, W., Arias-Navarrete, A., and Palacios-Pacheco, X. (2020). Proposal of an Architecture for the Integration of a Chatbot with Artificial Intelligence in a Smart Campus for the Improvement of Learning. *Sustainability*, *12*(4), 1500. https://doi.org/10.3390/SU12041500

Zhu, Q., and Luo, J. (2022). Generative Pre-Trained Transformer for Design Concept Generation: An Exploration. *Proceedings of the Design Society*, *2*, 1825-1834. https://doi.org/10.1017/pds.2022.185

Appendices

Following open access policies,[7] has been published:

1. The three texts produced by ChatGPT and the three texts produced by students.
2. The full instrument used in the study.

Date of receipt: 12 April 2023.
Review date: 15 May 2023.
Date of acceptance: 19 December 2023.

[7]Disponible en: https://osf.io/3xm98/?view_only=426e95fda4394a80b606874e91e25c82