And Plato met ChatGPT: an ethical reflection on the use of chatbots in scientific research writing, with a particular focus on the social sciences

Reyes Calderon (mrcalderon@comillas.edu) (universidad pontificia de comillas) and

Francisco Herrera

Abstract

This interdisciplinary paper analyzes the use of Large Language Models based chatbots (LLM-chatbots), with ChatGPT the most known exponent, in scientific research writing. By interacting with LLM-chatbots, researchers could reduce efforts and costs as well as improve efficiency, but taking important risks, limitations, and weaknesses, which could highly-order erosion scientific thought. While many scientific journals, as well as major publishers such as Springer-Nature or Taylor & Francis, are restricting its use, others advocate for its normalization. Debate focuses on two main questions: the possible authorship of LLM-chatbots, which is majority denied because their inability to meet the required standards; and the acceptance of hybrid articles (using LLM-chatbots).

Very recently, focusing on the education area, literature has found analogical similarities between some issues involved in Chatbots and that of Plato criticisms of writing, contained in the Phaedrus. However, the research area has been neglected. Combining philosophical and technological analysis, we explore Plato's myth of Theuth and Thamus, questioning if chatbots can improve science. From an interdisciplinary perspective, and according with Plato, we conclude LLM-chatbots cannot be considered as authors in a scientific context. Moreover, we offer some arguments and requirements to accept hybrid articles. We draw attention to the need for social science publishers, an area where conceptual hypotheses can take a long time to confirm, rather than solely on experimental observations. Finally, we advocate that publishers, communities, technical experts, and regulatory authorities collaborate to establish recommendations and best practices for chatbot use.

Keywords: LLMs based chatbots, ChatGPT; Authorship; Hybrid articles; Ethics on Research; Social Sciences.

Introduction

Artificial intelligence (AI) is rapidly transforming several aspects of our lives. We have joined the age of AI. Large Language models (LLMs) based chatbots (hereafter, LLM-chatbots), like ChatGPT, are the great exponent of the hype generated in this decade. LLM-chatbots are increasingly used in the domain of investigation and scientific publication (Giray et al, 2024). The fact that communication between scientists involves both researchers and chatbots (which are capable of producing new content, without the immediate supervision of humans) is something unique, unknown, and with very profound implications (Dwivedi et al, 2023). In fact, this tool is changing how science is done (Van Noorden, 2022).

By interacting with LLM-chatbots, researchers could reduce efforts and costs as well as improve efficiency (Cf. Table 2, Appendix), but taking important risks (Kendall & Da Silva, 2024). ChatGPT's rapid adoption presents a wide range of concerns, limitations, and weaknesses (Cf. Table 1, Appendix), which could highly erode scientific thought (Stokel-Walker, 2023). For instance, despite GPT-4 being technically a significant improvement over GPT-3.5, 18% of its citations are still fabricated and 24% contain significant errors.

Faced with this ambivalence, the scientific community is divided. While many journals and major publishers (Cf. Table 3, Appendix) are restricting its use, other publishers and journals consider that the prohibition is unfeasible and simply ask that its use be reported (Sallam, 2023). Debate focuses on two main questions (Thorp & Vinson, 2023; Lund & Naheem, 2024; Nazarovets & Teixeira da Silva, 2024): the possible authorship of chatbots; and the acceptance of hybrid articles (using LLM-chatbots).

Very recently, some authors have found the fears we see with LLM-chatbot echo, in some ways, the fears Plato captured in the Phaedrus (Bedington, et al, 2024), establishing

analogical similarities between Plato criticisms of writing and some issues involved in chatbots such as authenticity (Deptula et al, 2024), moral hazard (Loos & Radicke, 2024) or plagiarism (Misra & Ravindran, 2021). While most articles focus on the education area (Aylsworth & Castro, 2024; Bingham, 2024; Kitzinger, 2024), employing an instrumental analysis of pros and cons such as literature synthesis, citations, data analysis, etc., (Rahman et al, 2023), the research area has been neglected. Combining philosophical and technological analysis, we will explore Plato's myth of Theuth and Thamus as a means of providing insight and approaches to the use of ChatGPT on research, critically discussing LLM-chatbots authorship and hybrid articles in academic journals. We will conclude LLM-chatbots cannot be considered as authors in a scientific context. Moreover, we offer two practical and one theoretical argument to accept hybrid articles produced with the LLMs-chatbot's help, but with some preliminary requirements with clear specification of where and for what purpose an LLM-chatbot must be used. Finally, we advocate that publishers, communities, technical experts, and regulatory authorities collaborate to establish recommendations and best practices for chatbot use.

Today, LLM-chatbots could generate incorrect ("hallucinations") and/or inappropriate or unacceptable results (Alkaissi & McFarlane, 2023, Xu et al, 2024; Waldo and Boussard, 2024), but, after their evolution in our hands, some of these problems could be solved (Tonmoy et al, 2024). The real challenge lies not in unresolved issues or cumbersome functionalities, but in determining how and to what extent researchers should employ chatbots to create value for science and society. This requires critical thinking and understanding the potential hazards for science, beyond merely calculating costs and benefits.

We emphasize the need for social science publishers to be aware of the risks associated with data, algorithms, and hypotheses. This area requires special attention, as hypotheses

are often based on conceptual frameworks that take a long time to confirm, rather than solely on experimental observations.

To meet this objective, this paper is divided into eight sections. The second section briefly presents the methodological approach. The third section realizes a comprehensive and multidisciplinary review of the literature to investigate the nature and context of advanced LLM-chatbots in scientific research, describing both benefits and concerns. We make an aside to consider the case of social sciences, where literature detects a certain naivety when judging algorithms and data. The fourth section explores Plato's myth of Theuth and Thamus, obtaining critical resources for the formulation of a critical analysis focusing on scientific authorship and hybrid publications. Combining both philosophical and technological views, the fifth section critically analyzes what is an author? The sixth section analyzes the hybrid articles, closing with a formulation of requirements that align ChatGPT with the values of science. The seventh section shortly analyzes the specificity of the social sciences. Finally, we present our conclusions and describe the limitations of this study. An additional appendix includes three tables outlining a wide range of concerns: the limitations and weaknesses, the advantages of using LLM-chatbots for research as discussed in recent literature, and the analysis of the positions of leading journals and major publishers.

Methods: Methodological framework for a critical analysis

Given the research question, in formulating this article, we have incorporated several key scientific methods as a methodological framework. It is focused on two analyses. On the one hand, we employ a comprehensive and multidisciplinary literature review, analyzing opinions from various authors, updated with the latest advances. This review does not merely aim to effectively present relevant information on the research topic, which would allow researchers to become familiar with the relevant concepts and ideas. Our extensive

review attempts to help scientists build background knowledge so that they can bring an opinion. To achieve a holistic understanding of the topic, the article conducts a multidisciplinary exploration.

On the other hand, we tackle an analogical critical thinking based on the analogical discussion on Plato's criticisms of writing. Literature confirms its relevance in science and technology contexts (Cf. Kosar, 2024, p. 8), due to its capacity for (a) making sense of technological systems for lay people, (b) facilitating epistemological reflection on chatbots from a novel perspective, and (c) understanding and assessing emerging technologies in dialogue settings such as the scientific communities (Cf. Bucchi, M. & Trench, 2008). In contrast to scenarios of science fiction and permitting to know how society responded to previous technologies and the consequences these responses entailed, analogical thinking can enhance anticipatory capacities about governance processes (Schwarz-Plaschg, 2018).

We use analogous discussion on Plato's criticisms of writing, as a means of providing insight and approaches to the use of ChatGPT on research. With Clements (2022), we observe analogy between critics in the chatbots age and those that Plato observed, including the misuse or abuse of published information, the lack of context and, in the end, the role of memory. The Platonic distinction writing/logos; appearance/reality; verisimilitude/truth (Cf. Derrida, 1981, p. 181) permit us, arguing that, as a pharmakon, LLM-Chatbots could be a poison but also a medicine, which can be associated with better knowledge.

This methodological approach allows us to propose our hypothesis: An LLM-chatbot should not be considered an author of a paper, but it can serve as a tool that assists in the writing process, fitting into the hybrid article model. This hypothesis is explored in the

following four sections, guided by the methodological framework and the critical analysis conducted by the authors.

LLM-chatbots and science: a comprehensive analysis and literature review

Technically, LLM-chatbots are designed to understand and generate human-like text based on the input they receive. They integrate multiple technologies such as deep learning, unsupervised learning, instruction fine-tuning, multi-task learning, in-context learning and reinforcement learning (Cf. Wolfram, 2023). Its first pillar was the Transformer architecture, presented in the paper "Attention is all you need" by Google researchers (Vaswani et al, 2017), in which they claimed to have reached a new state of the art in generative AI. The experts agreed. In a macro survey conducted that year (Grace et al, 2018), they dated the years in which machine performance would surpass that of humans: translation (2024), high school essay writing (2026), truck driving (2027), writing best-sellers (2049) or general surgery (2053).

Based on a Transformer framework via unsupervised learning, in 2018, OpenAI, a non-profit organization with open-source software and publicly available advanced tools, presented the second pillar: the Generative Pre-trained Transformer model (GPT), a massive deep learning language model. The tool was trained on Microsoft Azure's AI supercomputer using 45 terabytes of text from Common Crawl, WebText2, books and Wikipedia (Brown et al, 2020). In the initial stage of generative pretraining, the algorithm assigns numerical values to words. Words that frequently appear together are associated within an "embedding," a multidimensional representation of their relationships. Chatbots utilize these embeddings to generalize concepts by leveraging the relationships between words in the training data. During this phase, weights are adjusted to minimize loss, which is the difference between the desired and actual results. The outcome is a tool capable of

generating functional "knowledge" across various contexts, which is then refined through reinforcement learning with human feedback, involving supervised fine-tuning on specific tasks relevant to the user.

ChatGPT has been iteratively updated from GPT-1 to GPT-4 (2018-2024) building through an easy-to-use web interface and capable of generating conversational-mode responses similar to those of an “expert” human to queries formulated in natural language, based on the input provided. GPT-4, with accepting image and text inputs and emitting text outputs, exhibits human-level performance on various professional and academic benchmarks (Wu et al, 2023). If the GPT parameters were increased tenfold in GPT-2 (one hundred times in GPT-3), the GPT-4 version has 500 times more parameters than GPT-3, approaching the number of neuronal connections in the human brain and proving to be functional for a wide range of applications (cf. Rudolph et al, 2023). OpenAI recently introduced a new series, designed to spend more time reasoning through problems before responding, which helps them solve more complex tasks in areas such as science, coding, and mathematics (Cf. Zong, et al, 2024). For instance, the GTP o3 LLM model represents a significant advancement in AI capabilities, particularly in complex tasks such as coding, mathematics, and science. It demonstrates improved performance and reasoning skills, making it a valuable tool for various applications. The model achieved a breakthrough high score of 87.5% on the ARC-AGI benchmark, which is designed to test genuine intelligence (human performance is benchmarked at 85 percent).

Along the way, OpenAI became a for-profit company, giving Microsoft exclusive access to its GPT-3 language model. The open became partially closed (the architectures for GPT-3 and GPT-4 have not been published), and the democratization of AI became doubtful (Metz & Weise, 2023). The issue is not trivial because, with huge economic and

environmental costs in terms of energy and water (Hao, 2020), innovation is being developed by only a small number of Big Tech players (United Nations, 2024).

While some authors emphasize an LLM-chatbot does not think (Dehouche, 2021), others emphasize it can pass public exams in medicine (Kung et al, 2023) or law (Katz et al, 2024), create poetry indistinguishable from human poetry (Köbis & Mossink, 2021) or display social intelligence at the level of licensed psychologists and doctors (Sufyan et al, 2024). Table 2 (appendix) summarizes main advantages for researchers, classified into efficiency improvement, quality improvement and cost saving in terms of time, effort, and money. However, risks and concerns described in Table 1 (appendix) are also present in scientific research, including incorrect and fictitious answers; fabricated or biased literature; inaccuracies in the code; limited logical reasoning skills; plagiarism; overconfidence; copyright, security or privacy violations, etc. To provide more clarity, we should have holistic performance metrics, considering both pros and cons. Currently, we lack reliable metrics.

However, a positive balance does not guarantee that its use will improve the quality of our scientific activity and our publications. We must remember chatbots are nondeterministic by nature. This is because they predict the probability of a word given the context, represented by a sample of words, which produces randomness (Ouyang et al, 2024). As a result, identical instructions/prompts can yield different responses to separate requests, and to check for hallucinations is not possible (Alkaissi & McFarlane 2023; Xu et al, 2024), which affects the reliability and reproducibility of research. Therefore, the potential of LLM-chatbots to improve (or depreciate) research must be carefully analyzed and constantly monitored.

The main examen is the evaluation of the hard changes in individual and collective memory. Let's not fool ourselves. As an algorithmic crystallization of memory, ChatGPT

can generate human-like texts from a simple message without further human intervention. But, far from being an AI agent that (re)distributes and (re)activates memories, ChatGPT is a kind of active guardian of memory, effectively deciding which sources of information and which interpretations of the collective past gain more visibility and are accessible to the platform users, and therefore shaping how this past is remembered or a decline in memory retention for humans (Bai et al, 2023).

As an entity responsible for the organization and retrieval of information in response to human queries, and given its non-deterministic nature, ChatGPT can change our scientific practices. In fact, our truth discovery becomes a challenge, which integrates multi-source noisy information by estimating the reliability of each source. For instance, the algorithmic initialization of the truths can be obtained using voting/averaging approaches, using weighted vote for categorical data and weighted median for continuous data to update truths (Cf. Li et al, 2016). Statistic could supplant truth, as, in the past, writing attempted to supplant logos. To expose this factor, we will turn to Plato.

Plato and ChatGPT: philosophical roots of the criticism

In ‘The Phaedrus’ (274e7–275b2), Plato relates the myth of the Egyptian King Thamus and the god Theuth, inventor of writing. Animated by the desire of popularizing his invention, Theuth presents it to Thamus, as "an elixir of memory and wisdom "(274 e), the best cure (pharmakon) for oblivion and ignorance. But the king argues that it is not an elixir of memory (mneme), but of reminding (hypomnesis) and can be a poison. While writing allows for a vast accumulation of knowledge, it can affect both (a) our ability to remember (it makes us forgetful), and (b) the integrity and power of the individual to interact with knowledge and reality.

The Thamus' critique is developed in three steps.

The first one is the criticism of technical manuals, which, according to Thamus, is needed but not sufficient condition for producing “results that are clear or certain”. Despite a set of true and useful theorems would be contained in a manual of mathematics, the owner of this manual is not a mathematician. A mathematician must be in possession of a set of abilities and skills, which permit to correctly understand propositions, to articulate the relevant true propositions to some concrete problem; to explains why they are relevant and true; to know how to justify decisions to others, as well as be able to teach others to become accomplished practitioners like himself. Those capabilities cannot be obtained in technical manuals (Cf. Rabbås, 2010, p. 32-33). Finally, the accumulation of data, writings or technical manuals, which can be deposited in external libraries or technological artifacts, by itself do not generate knowledge, and their memorization does not form a living memory. It is a “pharmakon-poison,” which connects with hypomnesic memory and makes us forgetful, while providing us with the illusion of wisdom.

The second step is developed across the analogy with painting. (275d4– e5). “Writing, Phaedrus... is very like painting; for the creatures of painting stand like living beings, but if one asks them a question, they preserve a solemn silence. And so it is with written words; you might think they spoke as if they had intelligence, but if you question them... they always say only one and the same thing" (275 de). According to Plato, the written word is dead because it cannot enter in a dialogue, adjusting to each person through a dialectical exchange, neither to specific audiences and situations. Writing is a “pharmakon-poison” (275e) for the memory because (a) it cannot answer the questions it raises; (b) it can reach people to whom it is not addressed, and it can be misinterpreted or misinterpreted; (c) when questioned, it cannot defend itself (Wieland, 1976; Barceló-Aspeitia & González-Varela, 2023).

In the third step, Thamus focuses on knowledge "in the soul of the student" (276a1-b1). Only the text capable of touching the soul of the student achieves the goal, which is to provide the resources that the student needs to support his practice toward truth. We are talking about a certain tacit knowledge (Polanyi, 1958), which is difficult to write or visualize, but which can make the writing alive for the student. In this way, the crucial difference between speaking in person and writing for others is not linguistic structure but efficacy as residing in truthfulness (Hyland, 1968, p. 39; Cf. Heidegger, 1931/32) and in readers able to be cojoined with true knowledge. If wisdom is not a manual but includes a set of abilities and skills to do things and search truth, and these abilities and skills are properties of the soul, not merely part of the content of his or her mind, then, when the discourse is successful planted, and, with time rooted and grow, the results change the learner (276e4–277a4). The scientist needs to remember occasions when she/he exercised his knowledge, which can be used as objects of comparison to instruct her/himself (or others) about what was right or best in that situation. On the opposite, the text without soul is as petrified, static and passive writing, dead words, being easy to remember, is only susceptible to identical repetition, without criticism or reinterpretation, dogmatic. However, since their appearance does not sufficiently evidence their instrumental character, they can give us the illusion of embodying knowledge.

After criticizing writing, Plato wrote dialogues. Why? According to literature (Cf. Clements, 2022; Rabbås, 2010; Staehler, 2013; Wieland, 1982), Plato's criticism of writing is not absolute, and, therefore, to interpreting writing as a pharmakon-remedy is possible.

The most obvious benefit of writing is to have a certain generative power of its own, which overcomes the limitation imposed by time and space (Kahn, 1981). This power could preserve and defend, for example, Socrates’s memory as well as the true concepts

to those who are unable to participate in dialectic (Allen, 2011, p.61). Another benefit is that writing text is always “on record” and accessible to scrutiny. As Derrida (1981, p. 113) shows, this element is present in Plato's perception of the need for the law to be written. In summary, the reader could use writing as remedy, recognizing certain conditions.

These conditions are described at the end of the Phaedrus. Socrates argues that good discourse has two features: their seeds (a) must have been planted in minds able to philosophy, and (b) must be cojoined with true knowledge (Cf. Allen, 2021 p. 60). When Phaedrus, to the detriment of philosophy, shows its preference for rhetoric which is cleverer, persuasive and requires less effort, Socrates argues that he never will be a good rhetorician if he does not become a philosopher. The aspiration of permanence and vividness is not only related to time and space, but also to the common (republic) good, which requires being based on truth (Zuckert, 2009, p. 323). Thus, the first crucial difference between speaking in person and writing for others is not linguistic structure but efficacy as residing in truthfulness (Hyland, 1968, p. 39; Cf. Heidegger, 1943). The second condition is in the abilities and soul of the reader, who must be intimately fused with genuine knowledge. In the writing law, this condition is satisfied because “if the writer is the legislator, the judge is its reader” (Derrida, 1981, p. 113).

A similar situation appears in science. Scientists are not only knowledge-gatherers or candidates to authorships, but also as societal stakeholders and as members of a scientific community, which protects a legacy (Mills & Sætra, 2024). As Polanyi (1962, p. 7) underlines, the activities of scientists take place around the published results of other scientists into the current professional standards of science, which must impose a framework of discipline and at the same time encourage rebellion against it. The authority of science over the lay public is held by a multitude of individuals, who controlling the

access of candidates until they reach a degree that allows them to assume the responsibility of joining that *republic*. In the end, writing as "Pharmakon-remedy" is embedded in the legacy, which can be collected, interpreted, criticized or discarded, a living memory correlated with the anamnestic memory, with which one converses, renews or adds new meanings, and is therefore non-dogmatic. While the "Pharmakon-poison" can equate appearance and truth causing the degradation of science (Hendrycks et al, 2023; Loos & Radicke, 2024), the "Pharmakon-remedy", closely connected with traditions and hermeneutic filters, helps to lead the person to the truth in perpetual tension between the temporal and the eternal, the real and the ideal, the finite and the infinite.

Therefore, the question here is if texts produced by LLM-chatbot must be considered dead letters, or a legacy with a soul. If, finally, a platonic dialogue between humans and chatbots is possible, chatbot should be considered an author; in other case, no. This question is important in practice because, despite humans, who are highly social beings, tend to respond to complex machines as if they were as "social" as we are (Sætra, 2020; cf. Foss & Saebi, 2017). This will allow us to steer the discussion towards our hypothesis of the use of LLM-Chatbots for hybrid articles.

Authorship. What is an author?

In "What is an author?" (1969), Foucault wonders whether a scientific article is not simply something written by a scientist and recalls that in the Middle Ages texts only had truth value if they were signed by their author. Two decades later, the same question is focused on a computer (Samuelson, 1990), and fifty years later, Ginsburg & Budiardjo (2019) reformulate the question focusing the most technologically advanced machines. They suggest machines are little more than faithful agents of the humans who design or use them, supporting the "right" question is how to evaluate the authorial claims of the humans involved in either preparing or using the machines that "create".

Nowadays, authorship is described since scientific standards (cf. ICMJE, 2023) which require substantial contributions to: a) the conception and design of the article; b)the collection, analysis and interpretation of data, c) drafting or revising the article critically; and d)final approval of the version to be published.

Although some argue that the idea of "substantial contribution" is too subjective and that there are important non-intellectual contributions that should not be unfairly neglected, chatbots cannot approve the final version, nor be responsible for all aspects of the work to ensure accuracy and completeness (Van Woudenberg, et al, 2024); nor understand a conflict-of-interest statement, hold copyright or have independent affiliation (Flanagin et al, 2023; Teixeira da Silva & Tsigaris, 2023).

According with Plato, there are other aspects which must be considered as author:

Firstly, "*the creation itself*". Beside intellectual property and the entitled to own, to grant permission to reproduce the final version, to appropriate the revenues and the academic and social prestige that publications generate (Efthyvoulou, 2008), authorship implies moral identity, whereby the researcher must recognize him/herself as a creative author (Epstein et al, 2023) and be competent to make judgments in a scientific community and take responsibility for them (Nannini, 2023; Scerbo, 2023). Authors are called to defend their work if it is challenged by scientific community, editing and critically review their contribution to science. Moreover, only an author can challenge legacy and the rules of the scientific method. Even if, in the future, chatbot could be conceptualized as a legal person to whom we attribute taxes, criminal liability, etc., moral identity is needed: an author must be able to a dialectic challenge with community, to claim misinterpreted or misunderstood; and, when challenged, to defend.

Not everyone agrees. While countries such as Korea define a "work" as "a creation that expresses the thoughts or feelings of a human being" and an "author" as "a person who

creates a work", preventing chatbots authorship, India or Canada have already accepted applications for copyright registration with computer applications (Lee, 2023). Moreover, some authors reclaim to describing new standards and guidelines. For example, Polonsky & Rotman (2023) point to (a) articles written by 100 or more authors, with each author's contribution blurred; (b) articles signed by authors who died before publication; or (c) organizational authors with no identifiable individuals, with each author's scope of formal approval and responsibility like that of a chatbot. They also indicate that protocols such as Vancouver recognize that some authors have specialized roles and journals provide a list of author roles in which ChatGPT could be included. On the other hand, they stress that chatbots are becoming so sophisticated so quickly that they will be hard to ignore.

Secondly, "*the responsibility*". As chatbots lie beyond the boundaries where moral values apply, publishers can only demand from named individuals the basic requirements of honesty, ethics and integrity that are essential for journals to advance science. Without proper attribution of authorship, it is unclear who is responsible for the content of the manuscript, the review, the veracity of the sources, the unduplicated publication, or the inclusion of all relevant points of view. (Ginsburg & Budiardjo, 2019). Breaches of codes of conduct and ethical behavior in scientific publishing (certainly more common in the social sciences according to (Xie et al, 2021)) undermine the integrity of the whole system, damaging both the reputation of journals and the value of science. Whether through misconduct (falsification, fabrication, plagiarism) or questionable research practices (cf. Xie et al, 2021), only individuals can be held accountable. Editors must therefore certify that authors are who they claim to be and that they have participated sufficiently in the study and critical analysis to be able to publicly guarantee its content and adherence to the rules of the scientific method and to guarantee the intellectual property of the ideas disclosed.

Third, "*the membership*". The scientist is always, first of all, a reader/student, who receives the written past, appropriates it and unravels for himself the meaning it suggests to him. Before this appropriation, he must ask himself where the information comes from, and how much truth it contains. Because the context in which it was written differs from that of researcher, author requires clarifications to illuminate the truth of the message (Staehler, 2013, p. 87). Some of these come from knowing the author of the quotes because, through his career, we can get closer to his thoughts and, as Foucault pointed out, evaluate their reliability. As for the choice of one method or another, this "truth" is even more diffuse. Placing a text within a particular topic legacy is a quality of an expert specialist (Gupta, 2024). An expert is someone who possesses extensive and authoritative knowledge in a particular area that is not possessed by most individuals in a community, being candidate to trust (cf. Croce, 2019). Because a chatbot cannot distinguish between "accurate and false information" (Bhattacharyya et al, 2023, p. 6), the success criterion of an LLM-chatbot is verisimilitude, not veracity, being not candidate to trust. Only those who have a thorough knowledge of the scientific literature and the various theories on a topic can place hypotheses in a scientific context, distinguish meaningful connections from seemingly irrelevant information, make accurate deductions, and apply this knowledge to specific situations. A chatbot does not have the capacity to perform this task, which inevitably requires a degree of human judgment.

The three conditions are resolved in scientific publication through the dialectic between authors and referees, and among authors into the scientific community. This dialectic would be the remedy, or the "antidote" to poison (Cf. Horbach & Halffman, 2020).

A special emphasis requires the role of editors and reviewers. If a journal is the dynamic outcome of a collective and interdisciplinary effort among scientists to create an ethical, open and shared framework for conducting scientific research. The role of the editor is to

ensure that the framework remains open, thereby facilitating critical analysis and the search for truth inherent in any scientific endeavor. By validating any contemporary research in a competent journal, editors possess a valuable and human cooperative tool to disagree and move forward. An editor who is responsible for aligning algorithmic thinking that can benefit certain stages of the research program with ethical values in scientific production must have a clear understanding of this issue.

Some recent works support the use of LLMs as expert reviewers (Lu et al, 2024), permitting fast and lower-cost evaluations and decreasing the number of needed expert reviewers (Cf. Baek et al, 2024; Lu et al, 2024). Analyzing the use of LLM-chatbots for peer review, Donker (2023) finds good summaries, description of the main aim and conclusions, comments on style and some constructive criticism, but also hallucinations, incorrect or uncorrelated answers to the text or not appropriate to the data, recommendation of non-existent articles and intellectual property problems. He concludes that reviewers should refrain from using chatbots tools and be trained to detect when and how a chatbot has been used correctly.

According to the Aristotelian tradition, the crucial function of a journal is not epistemic (i.e., facilitating "understanding of the underlying principles of the world," where a chatbot, despite its numerous shortcomings, can be of great help), but to assist in "scientific performances in the world," which is a matter of phronesis, or practical wisdom (Cf. Lee, 2005). This is a dialogical relationship between thought and action (Atkin, 2007, p. 68), where a chatbot is not effective. Considering only the epistemic function of AI, it may partially displace human (Cohen, 2013, p. 1926). However, when it comes to practical wisdom, an artificial science without critical thinking (Cohen, 2013, p. 2921) and without room for altruism, empathy, and other essential elements of human

flourishing does not be accepted. We understand that neither the accumulation of data nor speed can override academic analysis and the critical search for truth.

But if the editor reviewing the text wants to be sure and asks why a particular work by a particular author is cited or not, or why a particular method is used, the editor will not get an answer: a chatbot is not able to explain or justify its choices (cf. Chavanayarn, 2023; Thorp & Vinson, 2023; Stokel-Walker, 2023).

In summary, while ChatGPT passes the Turing test and convinces people that its answers are human (Guo et al, 2023), produces academic texts that are difficult to distinguish from those written by humans (Campbell et al, 2022), generates high-quality, evidence-based research questions (Lahat et al, 2023), and writes a convincing manuscript in less than an hour (Elbadawi et al, 2024), it does not possess the ability to distinguish between accurate and false information. The responsibility for the creation and verification of information inevitably requires human judgment. Therefore, its outputs are not considered scientific documents, and it cannot be regarded as a scientific author of a text or review.

Hybrid articles: The path to help with LLM-Chatbots

A scientist working with a chatbot produces a polyphonic response (Kjeldsen, 2024), which contains the voice of the scientist in a choral scientistic music, and another (or several) voice of the machine. Last one, clearly distinct from the first, transmits (at best) no-contextualized messages attributable to other human beings who are not subject to current scientific authority (Sætra, 2020) and mission.

Socrates distinguishes between the writer who is merely “a maker of speeches", whose mission is to persuade the audience (277e5–278b4), and the writer who “has composed his writings with knowledge of the truth” (276d-78e). Criticizing the first, Socrates signals that, without truth, students will become wise in their own opinion (doxa) instead of wise (sophos) (275a–b), opposing the probable (eikos) (that "happens to spring up in

the many, through likeness versus the truth") to the truthfulness, and finally the "public" opinion (doxa) to knowledge (episteme) of scientists (246A-49D). While scientists contribute episteme to the polyphonic sound, chatbot contributes doxa based on eikos.

For poorly prepared scientists or those outside of science, the outputs of a chatbot will create a kind of illusory knowledge. If it were false or erroneous, they will not identify it and will report it as true. Learning without teachers (Phaedrus 275a) is equivalent to doing science without scientific contrast: it can produce the illusion of wisdom in the ignorant, but it cannot expand knowledge.

However, a good scientist is able to produce a harmonic song with human and machine songs. We have two practical arguments and one theoretical one for accepting hybrid articles produced with the help of the LLM-chatbot.

The first one is the conviction that AI deployment is unstoppable and, therefore, a policy of non-co-production unfeasible in the long term. Recent studies reveal that two-thirds of authors are utilizing chatbots without disclosing this fact (cf. Khalifa et al, 2024; Lund & Naheem, 2024; Alkaissi & McFarlane, 2023), creating fairness and legality problems for publishers. Assuming that authors are being honest when stating that they do not use them or when not stating that they use them is an unviable option.

The second one is that currently it was extraordinarily challenging to detect what articles have been produced with an LLM-chatbots (Gao et al, 2022). Despite to essays (cf. Van Dis et al, 2023; Köbis & Mossink, 2021), there is not motivation to invest in this function, appearing as unprovable to have soon have effective detectors. In their own self-interest and given the irreversibility of the publication decision and its implications (Stahl et al, 2022), publishers have strong incentives to reject (Gordijn & Have, 2023) or to reclaim transparency.

Finally, on the theoretical argument, there is a more compelling reason. As human agents, researchers intend both the end and the means, praxis and poiesis. On the one hand, they intend the improvement of the science itself, with centrality on persons and truth. On the other, they intend its practical development, measured, for example, in investigation's results (i.e., papers, patents, etc.), which are influenced by something external (regulations, ethical guides, journal requirements, etc.) elements and are executed only to the extent necessary to achieve. This last has embedded some poietic activity. Despite, by nature, chatbots are unable to praxis, which implies logos and the creation of a shared reality, chatbots can emulate many of the poietic elements embedded in scientific activity. This could suppose a major but risky advancement in specific scientistic tasks, which must be clearly described. Here, it is crucial to develop a protocol, which should clearly describe the whole analysis process, which at the same time, can be useful to the realm of creativity and efficiency.

In summary, we advocate avoiding restrictions, but we call for a unifying comprehensive ethical framework for their use, in terms of advertising transparency, accountability or responsibility on the use of chatbots, and authorship, as well as for publishers/editors.

Advertising transparency recommendations range from those asking for basic information to be attached in the acknowledgments section, to those asking for strong details of (a) its specifications (name, version, model and source), (b) how it has been used and (c) why, and categorizing it in the "Methodology" sections. Based on the literature and our own judgment, we agree with the stronger version. With (a) because different chatbot models exhibit varying types and degrees of biases, employ diverse workarounds, and possess distinct characteristics (cf. Sufyan et al, 2024), with (b) because a chatbot can be used in many ways and for many purposes, with different effects (cf. Zhuo et al, 2023) and with (c) because it is necessary to know who is "playing that

polyphonic tune" to maintain trust in science. We agree with the proposal of Jenkins & Lin (2023), who ask journals to provide procedures to give transparency about how and what content has been generated with AI and, mainly, for what purpose.

We also require authors to taking individual and/or collective *formal responsibility* for all contributions made using chatbots, including the accuracy and proper attribution of all cited material, and to search for and cite omitted sources (cf. Taylor, 2024). This is in order to avoid liability gaps, which in the case of scientific articles is usually a forward-looking liability that is more focused on the actor than on the act itself (cf. Hedlund, 2022).

Third, *on the authorship*, we add an additional condition to "satisfy Plato". Since editors are responsible for ensuring that authors answer the questions raised by their text and that it is correctly interpreted, we recommend authors defend their text and describe what their contribution to science has been.

The particular focus and specificity of the social sciences

Some authors argue that chatbots could help especially social sciences, improving the objectivity of the scientific field to examine practicing scientists' views of nature of science and explore possible relationships between these views and science social context. This offering a less arbitrary epistemology, purging the contamination of individual reasoning and improving the accuracy of results both at the initiation stage of research and in data analysis, reducing the potential for human error and achieving better reproducibility (Burger et al, 2023).

However, other authors warn against a certain naivety in social scientists who identify themselves as "end users", discussing the many limitations of LLM-chatbots. Bails examines how bias in the data used to train these tools can negatively impact social

science research—as well as a range of other challenges related to ethics, replication, environmental impact, and the proliferation of low-quality research (Bail, 2024).

A LLM-Chatbot is targeted at technologically literate users. To use it effectively, social researchers must be willing to fully embrace that technology, which often involves collaborating with experts, fostering interdisciplinary collaboration, and investing in training and development programs. Faced with a technology that causes radical changes, a scientist must not only know how to use it, but also gain the ability to understand and critique its results and use and to manage correctly the data. However, many social scientists are unaware of its risks. We want to draw attention to two gaps that a non-expert social scientist might face managing data.

The first gap stems from its theoretical burden. Data are not neutral "facts", value-free observations, but imply goals, means, constraints and selection criteria (Sarasvathy 2001, p. 249). As masterfully Coveney et al (2016) put it, big data need big theory too. As S. Javed et al (2021) pointed out, "letting nature speak for itself" is not the same as "letting the model speak for itself", as the latter is based on a priori assumptions and theories.

The second gap originates from the data selected for training. Fossil fuels may be polluting, but they produce the same result (heat) as clean energy. This is not the case here: the performance of LLM-Chatbots depends directly on the quality and relevance of the data on which it is trained (cf. Dwivedi et al, 2023). There is evidence that tools trained on problematic data reproduce problematic associations (Grassini, 2023) in the form of stereotypes (Font & Costa-Jussa, 2019), abusive language and hate speech (Kennedy et al, 2018), and all kinds of biases of gender (Tschopp et al, 2023), race (Buolamwini & Gebru, 2018), sexual orientation or religion (Kurita et al, 2019), etc.

To ensure that technology is harnessed for the benefit of humanity and scientific progress without impeding valuable research, social science editors must possess skills in data

analysis, algorithm and code design, and training in the latest technological advancements. Given the scarcity of individuals who are both technologically proficient and ethically informed, it is crucial for journals to form interdisciplinary teams and seek advice from independent experts. This approach will help navigate the complexities of the new technological landscape effectively.

The first gap stems from its theoretical burden, which requires a deep understanding of the underlying principles and frameworks guiding technological advancements. The second gap originates from the data selected for training, highlighting the importance of using diverse and representative datasets to ensure fairness and accuracy in technological applications. By addressing these gaps, social science editors can better manage the integration of technology in research and publication, ultimately fostering a more ethical and progressive scientific community.

Therefore, in social science, the application of ChatGPT must be accompanied by a meticulous analysis of the data and its inherent characteristics. Equally important is the human contribution in crafting the conceptual framework of the paper, developing hypotheses, and establishing expectations for practical scenarios. As we have mentioned, the conceptual frameworks and hypotheses may take long time to confirm, that is, several years to validate. Therefore, it is imprudent to rely solely on the probabilistic nature of ChatGPT, which is grounded in text and learned knowledge, to generate hypotheses or conceptualize theories in the social sciences. This is in addition to everything indicated for hybrid articles, which is generic in nature in any scientific setting.

This circumstance makes us pay special attention to the particular case of the social sciences. In any case and scenario, any hypothesis for the future, taking a long time to confirm, in any field, must be proposed and validated by a human being. Judea Pearl's theory on imagination highlights a current challenge that remains beyond the capabilities

of LLM-Chatbots. Pearl's "Ladder of Causation" (Pearl & Mackenzie, 2018) emphasizes that while AI can observe and perform tasks, the ability to imagine and hypothesize about future scenarios is uniquely human. This imaginative capability is crucial for developing and validating hypotheses, particularly in the social sciences, where conceptual frameworks often require extensive time for confirmation. Thus, the human role in proposing and validating hypotheses remains indispensable, underscoring the limitations of AI in this domain.

Concluding remarks, limitations, and future directions

LLM-chatbot technologies are increasingly used in the domain of investigation and scientific publication (Giray et al, 2024). The fact that communication between scientists involves both researchers and chatbot (which can produce new content, without the immediate supervision of humans) is something unique, unknown and with very profound implications (Dwivedi et al, 2023). The huge benefits in terms of reducing efforts and costs as well as improve efficiency are measured against concerns, limitations and weaknesses, an evaluation still pending (Grassini, 2023; Stokel-Walker, 2023; Zhou et al, 2023).

LLM-chatbot passes the Turing test and makes people believe that its answers are human (Guo et al, 2023); produces academic texts that are difficult to distinguish from those produced by humans (Campbell et al, 2022); generates high-quality, evidence-based research questions (Lahat et al, 2023); or writes a convincing manuscript in less than an hour (Elbadawi et al, 2024), but it presents a wide range of concerns, limitations, and weaknesses (Stokel-Walker, 2023).

However, in our opinion, the main question is not about the balance between pros and cons, but if the integration of LLM-based chatbots in scientific writing improves or erodes scientific thought the quality and integrity of published research (Nazarovets & Teixeira

da Silva, 2024). The scientific community is divided. While many journals and major publishers (such as Springer-Nature, Elsevier, Lancet or Taylor & Francis) have restricted its use, other publishers and journals consider that the prohibition is unfeasible and simply ask that its use be reported (Sallam, 2023). Debate focuses on two main questions highlighted in the introduction: the possible authorship of LLM-chatbots, and the acceptance of hybrid articles (using LLM-chatbots) (Thorp & Vinson, 2023; Lund & Naheem, 2024; Nazarovets & Teixeira da Silva, 2024).

To contribute to the debate, we turn to Plato's criticisms of writing, contained in the Phaedrus, an analogy which present many connections with chatbots. Very recently, it has been employed in the analysis of some peculiar aspects (Deptula et al, 2024; Loos & Radicke, 2024) mainly focusing on education area and from instrumental analysis (Aylsworth & Castro, 2024; Bingham, 2024; Kitzinger, 2024; Rahman et al, 2023). However, research area and the analysis of science itself have been neglected. We explore Plato's myth of Theuth & Thamus as a means of providing insight and approaches to qualify the use of LLM-chatbots on research.

From an interdisciplinary perspective, our conclusions are as follows:

Firstly, it is necessary and urgent to establish a unified and comprehensive ethical framework for the use of LLM-chatbots in research. Editors, most of whom could suffer from a certain technical illiteracy, need that the publishers, communities, technical experts and regulatory authorities sit together and lay down recommendations and good practices about its use, an uncompleted mission jet.

Secondly. an LLM-chatbot cannot be considered as authors/co-authors or be cited as authors/co-author in a scientific context. Not only because they do not comply with the present editorial standards, but because they are unable to escape Platonic criticism of writing.

Third, hybrid articles may be acceptable under certain strong conditions. We offer some preliminary recommendations to help guideline: *Advertising transparency. formal responsibility* for all contributions made using chatbots, and *on the authorship*, we add an additional condition to "satisfy Plato", *the creation itself and the membership as author*, requiring clarifications to illuminate the truth of the message. A chatbot does not have the capacity to perform this task, which inevitably requires a degree of human judgment.

This paper provides a platform for debate, analysis, and reflection on the use of chatbots in the final writing of scientific papers. It contributes to the ongoing discussion within the scientific community, as cited in this paper. We hope it adds to the scientific legacy. As with any study, we acknowledge certain limitations, which also present opportunities for new avenues of exploration. We discuss some of these limitations.

Firstly, and due to the recentness, dynamic and novelty of the technology, the findings of this research seem to raise more questions than they provide answers. Advances in the reasoning processes of the new LLM-Chatbots models are ongoing, although the limitation posed by Judea Pearl remains on the horizon, imagine.

Second, very recently, OpenAI has launched the mentioned OpenAI o3, known as Strawberry, the third model in a new series designed to perform complex reasoning tasks, allowing it to generate more complex answers. precise and thoughtful. Despite we think our conclusions are also valid, this paper does not analyze this innovation. It would be interesting to analyze the use of the LLM-chatbots in the generalization of theories and hypotheses based on past data they have learned and on the reasoning capacity of the new models (Strawberry, Gemini, Llama family, Claude, etc.) (Movva et al, 2024). Advances in the complex reasoning processes of the new LLM-Chatbots models are ongoing.

Thirdly, an analysis of the impact of chatbots on the social sciences, and the responsibility question would require much more space than these pages allow, and a space-temporal analysis.

ACKNOWLEDGMENTS

We thank the prof. Raúl González Fabre for comments. We don't thank CHATGPT or other chatbot because this paper was totally produced by humans. This work has been partially supported by Knowledge Generation Project PID2023-150070NB-I00, funded by the Ministry of Science, Innovation, and Universities of Spain.

Appendix

TABLE 1, 2 & 3 HERE

References

Alkaissi H, McFarlane S. (2023) Artificial hallucinations in ChatGPT: Implications in scientific writing. *Cureus,* 15(2), 1–4.

Allen, D. (2011) *Why Plato Wrote*. John Wiley & Sons.

Atkin, J. (2007) What role for the humanities in science education research? *Studies in Science Education*, 43: 62-87

Aylsworth, T., & Castro, C. (2024) Should I Use ChatGPT to Write My Papers? *Philosophy & Technology*, *37*(4), 117.

Baek J. et al (2024) Research Agent: Iterative Research Idea Generation over Scientific Literature with Large Language Models. *ArXiv,* abs/2404.07738, 2024.

Bai, L., Liu, X., & Su, J. (2023) ChatGPT: The cognitive effects on learning and memory. *Brain-X*, *1*(3), e30.

Bedington, A. et al (2024) Writing with generative AI and human-machine teaming: Insights and recommendations from faculty and students. *Computers & Composition*, *71*, 102833.

Bhattacharyya, M. et al (2023) High rates of fabricated and inaccurate references in ChatGPT-generated medical content. *Cureus* **15**, e39238.

Bail, C. (2024) Can Generative AI improve social science? *Proceedings of the National Academy of Sciences*, *121*(21), e2314021121.

Barceló-Aspeitia, A. & González-Varela, E. (2023) Plato on False Judgment in the Theaetetus. *Journal of the History of Philosophy*, *61*(3), 349-372.

Bingham, C. (2024) Education and Artificial Intelligence at the Scene of Writing: A Derridean Consideration. *Futurity Philosophy*, *3*(4), 34-46.

Brännström, A., Wester, J., & Nieves, J. C. (2024) A formal understanding of computational empathy in interactive agents. *Cognitive Systems Research*, *85*, 101203.

Brown, T. et al (2020) Language models are few-shot learners. *Advances in Neural Inf. Proc. Systems*, 33, 1877-1901.

Bucchi, M. & Trench, B. (eds) (2008) *Handbook of public communication of science and technology.* Routledge, London.

Burger, B. et al (2023) On the use of AI-based tools like ChatGPT to support management research. *European Journal of Innovation Management*, 26(7), 233-241.

Buolamwini, J., & Gebru, T. (2018) Gender shades: Intersectional accuracy disparities in commercial gender classification, In *Conference on fairness, accountability & transparency:* 77-91

Campbell, C. et al (2022) Preparing for an era of deepfakes and AI-generated ads: A framework for understanding responses to manipulated advertising. *Journal of Advertising*, 51(1), 22–38.

Clements, E. (2022) Theuth, Thamus & digital civics: Plato's formulation of memory and its lessons for civic life in the digital age, *Memory Studies* 15.4: 767-783.

Chavanayarn, S. (2023) Navigating Ethical Complexities Through Epistemological Analysis of ChatGPT. *Bulletin of Science, Technology & Society 43*(3-4), 105-114.

Cohen, J. (2013) What privacy is for. *Harvard Law Review, 126*(7), 1904–1933.

Coveney, P.; Dougherty, E. & Highfield, R. (2016) Big data need big theory too. *Philos. Trans. of Royal Soc. A: 374*(2080), 20160153.

Croce, M. (2019) On what it takes to be an expert. *The Philos. Quart*, *69*(274), 1-21.

Dehouche, N. (2021) Plagiarism in the age of massive generative pre-trained transformers (GPT-3), *Ethics in Sc. & Environment Politics*, 21, 17-23

Deptula, A., Hunter, P., & Johnson-Sheehan, R. (2024) Rhetorics of Authenticity: Ethics, Ethos, & Artificial Intelligence. *J. of Business & Technical Communication*, 10506519241280639.

Derrida, J. (1981) *Plato's pharmacy*. In: *Dissemination*. The Athlone Press, London.

Donker, T. (2023) The dangers of using large language models for peer review; *The Lancet* 23 (7): 781.

Dwivedi, Y. et el. (2023) So, what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *Int. J. of Inf. Management*, 71, 1–63.

Efthyvoulou, G. (2008) Alphabet economics: The link between names and reputation. *J. of Socioeconomics 37*(3), 1266-1285.

Elbadawi, M. et al (2024) The role of artificial intelligence in generating original scientific research, *Int. J. of Pharmaceutics*, 652, 123741.

Else, H. (2023) Abstracts written by ChatGPT fool scientists. *Nature, 613*: 423–423.

Epstein, Z. et al (2023) Art and the science of generative AI. *Science*, 380, 1110–1.

Flanagin, A. et al (2023) Nonhuman "authors" and implications for the integrity of scientific publication and medical knowledge, *JAMA*, 329(8):637–639.

Font, J. & Costa-Jussa, M. (2019) Equalizing gender biases in neural machine translation with word embeddings techniques, *arXiv:1901.03116*.

Foucault M. (1969) *Qu'est ce qu'un auteur? in Dits et écrits*, I, Paris, Gallimard, 1994: 789-821,

Foss, N., & Saebi, T. (2017) Fifteen years of research on business model innovation: How far have we come, and where should we go? *Journal of management*, *43*(1), 200-227.

Gao, C. et al (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.*

Giray, L. et al (2024) Strengths, Weaknesses, Opportunities, and Threats of Using ChatGPT in Scientific Research, *International Journal of Technology in Education*, *7*(1), 40-58.

Ginsburg, J. & Budiardjo, L. (2019) Authors and Machines, *Berkeley Technology Law Journal* 34:343.

Gordijn, B. & Have, H. (2023) ChatGPT: evolution or revolution? *Medicine, Health Care and Philosophy*, *26*(1), 1-2.

Grace, K. et al (2018) Viewpoint: When will AI exceed human performance? Evidence from AI experts. *Journal of Artificial Intelligence Research*, 729–754.

Grassini, S. (2023) Shaping the future of education: exploring the potential and consequences of AI and ChatGPT in educational settings. *Education Sciences*, *13*(7): 692

Gupta, B. (2024) Can Artificial Intelligence Only be a Helper Writer for Science? *Science Insights*, *44*(1), 1221-1227.

Kjeldsen, B. (2024) Ethos in the Machine—The Rhetorical Character of Debate AI, in *Ethos, Technology, & AI in Contemporary Society: 10-24.*

Hao, K. (2020) OpenAI is giving Microsoft exclusive access to its GPT-3 language model. *MIT Technology Review. Retrieved September*, *25*.

Hedlund, M. (2022) Distribution of forward-looking responsibility in the EU process on AI regulation. *Frontiers in Human Dynamics*, *4*, 703510.

Heidegger, M. (1943) On the essence of truth. *Engagements Across Philosophical Traditions*, 243.

Hendrycks, D., Mazeika, M. & Woodside, T. (2023) An Overview of Catastrophic AI Risks. *arXiv:2306.12001*.

Horbach, S.P. & Halffman, W. (2020) Journal peer review and editorial evaluation: cautious innovator or sleepy giant? *Minerva*, 58 (2): 139-161.

Hyland, D. (1968) Why Plato wrote dialogues. *Philosophy & Rhetoric*, 38-50.

ICMJE (2020) Defining the role of authors and contributors. https://www.icmje.org/recommendations/

Javed, S. et al (2021) Understanding the role of objectivity in machine learning and research evaluation. *Philosophies*, *6*(1), 22.

Jenkins, R., & Lin, P. (2023) AI-Assisted authorship: How to assign credit in synthetic scholarship. *SSRN 4342909*.

Katz, D., et al (2024) GPT-4 passes the bar exam. *Philosophical Transactions of the Royal Society A, 382*(2270), 20230254.

Kendall, G. & da Silva, J. (2024) Risks of abuse of large language models, like ChatGPT, in scientific publishing: Authorship, predatory publishing, and paper mills. *Learned Publishing 37*(1), 55-62.

Kennedy, M. et al (2018) A typology and coding manual for the study of hate-based rhetoric. *PsyArXiv*, July: 18.

Khalifa, A. & Ibrahim, M. (2024), Artificial intelligence (AI) and ChatGPT involvement in scientific and medical writing, a new concern for researchers. A scoping review, *Arab Gulf Journal of Scientific Research*.

King, M. (2023) The future of AI in medicine: A perspective from a Chatbot. *Annals of Biomedical Engineering*, 51, 291- 295.

Kitzinger, C. (2024) OpenAI's Pharmacy? On the Phaedrus Analogy for Large Language Models. *Critical AI*, 2(1)

Köbis, N. & Mossink, L. (2021) Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry. *Computers in Human Behavior*, 114, 106553.

Kosar, A. (2024) Nietzschean Language Models and Philosophical Chatbots: Outline of a Critique of AI, *The Agonist* 18(1): 7 – 17.

Kung, T., et al (2023) Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models, *PLoS digital health*, *2*(2), e0000198.

Kurita, K. et al (2019) Measuring Bias in Contextualized Word Representations. In *Proc. of the First Workshop on Gender Bias in Natural Language Processing*. 166–172.

Lahat, A. et al (2023) Evaluating the use of large language model in identifying top research questions in gastroenterology. *Scientific reports*, *13*(1), 4164.

Lee, M. (2005) *Epistemology after Protagoras: responses to relativism in Plato, Aristotle, and Democritus*, Oxford, OUP.

Lee, J. (2023) Can an artificial intelligence chatbot be the author of a scholarly article? *Sci Edit* 10(1):7-12.

Leung, L.; Chu K. & Qiao M. (2021) AI literacy: definition, teaching, evaluation and ethical issues. *Proc Assoc Inform Sci Technol* 58(1):504–509.

Loos, E. & Radicke, J. (2024) Using ChatGPT-3 as a writing tool: an educational assistant or a moral hazard? Current ChatGPT-3 media representations compared to Plato's critical stance on writing in Phaedrus, *AI &Ethics:* 1-14.

Lu, C. et al (2024) The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery, ArXiv, abs/2408.06292.

Lund, B. & Naheem, K. (2024) Can ChatGPT be an author? A study of artificial intelligence authorship policies in top academic journals. *Learn. Publish*, *37*(1): 13-21.

Metz, C., & Weise, K. (2023) Microsoft bets big on the creator of ChatGPT in race to dominate AI. *The New York Times*, *12*.

Mills, S. & Sætra, H.S. (2024) Algorithms in the Room: AI, Representation, and Decisions about Sustainable Futures. Available at SSRN: 4952529.

Movva, R. et al (2024), Topics, Authors, and Institutions in Large Language Model Research: Trends from 17K, *arXiv*:2307.10700v3.

Nannini, L. (2023) Voluminous yet Vacuous? Semantic Capital in an Age of Large Language Models. *arXiv:2306.01773*.

Nazarovets, S., Teixeira da Silva, J. (2024) ChatGPT as an "author": Bibliometric analysis to assess the validity of authorship. *Accountability in Research*, 1–11.

Ouyang, T. et al (2024) Stability Analysis of ChatGPT-based Sentiment Analysis in AI Quality Assurance. *arXiv:2401.07441*.

Plato. *Phaedrus*. R. Waterfield traductor. OUP Oxford, 2002.

Pearl, J. Mackenzie D. (2018) *The book of why: the new science of cause and effect*. Basic books.

Polanyi, K. (1962) The Republic of science, *Minerva*, 1: 54-73

Polanyi, M. (1958) *Personal knowledge: Towards a post-critical philosophy*. London: Routledge & Kegan Paul.

Polonsky, M. , Rotman, J. (2023) Should artificial intelligent agents be your co-author? Arguments in favor, informed by ChatGPT. *Austr. Mark. Journal*, 31(2): 91–96.

Rabbås, Ø. (2010) Writing, memory and wisdom: the critique of writing in the Phaedrus. *Symbolae Osloenses* 84(1): 26–48.

Rahman, M. et al (2023) ChatGPT and Academic Research: A Review and Recommendations Based on Practical Examples. *Journal of Education, Management and Development Studies*. 3(1) 1-12.

Rudolph, J, Tan, S. (2023) War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *Journal of Applied Learning and Teaching*, *6*(1) 364-389.

Sætra H.S. (2020) The parasitic nature of social AI: Sharing minds with the mindless, *Integrative Psychological & Behavioral Science*, 54: 308-326.

Sallam, M. (2023) ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns, *Healthcare* 11(6):887.

Samuelson P (1990) Can a computer be an author? In Ermann, M. Ed. *Computers, ethics, & society*. (.): 299-307.

Sarasvathy S (2021) Even-if: Sufficient, yet unnecessary conditions for worldmaking. *Organization Theory 2*(2):26317877211005785.

Scerbo M (2023) Can Artificial Intelligence Be My Coauthor? *Sim. in Health.* 75: 215–218.

Staehler T (2013) Theuth versus Thamus: the esoteric Plato revisited, *Journal of Ancient Philosophy* 7(1): 65–94.

Stahl B et al (2022) Organizational responses to the ethical issues of artificial intelligence. *AI & Society 37*(1):23-37.

Stokel-Walker C (2023) ChatGPT listed as author on research papers: many scientists disapprove. *Nature* 613(7945):620–621

Sufyan N et al (2024) Artificial intelligence and social intelligence: preliminary comparison study between AI models and psychologists. *Frontiers in Psychology*. *15*:1353022.

Teixeira da Silva J, Tsigaris P (2023) Human- and AI-based authorship: Principles and ethics. *Learned Publishing* 36(3): 453– 462.

Tonmoy S et al (2024) A comprehensive survey of hallucination mitigation techniques in large language models. *arXiv:2401.01313*.

Thorp H, Vinson V (2023) ChatGPT is fun, but not an author. *Science* 379(6630):313.

Tschopp M, Salam H (2023) Spot on SDG 5: Addressing Gender (In-)equality within and with AI. *Technology and Sustainable Development:* 109-126.

United Nations. AI Advisory Body (2024) Governing AI for Humanity. Final Report September.

Van Dis E et al (2023) ChatGPT: Five priorities for research. *Nature* 614 (7947):224–226.

Van Noorden R (2022) How language-generation AIs could transform science. *Nature* 605:21–21.

Van Woudenberg R, Ranalli C, Bracker D (2024) Authorship and ChatGPT: A conservative view. *Philosophy & Technology*, *37*(1):34.

Vaswani A et al (2017) Attention is all you need. *Advances in neural information processing systems*, *30*.

Waldo J, Boussard S (2024) GPTs and Hallucination: Why do large language models hallucinate? *Queue* 22(1):19-33.

Wieland W (1976) Platon und der Nutzen der Idee. Zur Funktion der Idee des Guten. *Allgemeine Zeitschrift für Philosophie* 1(1):19-33.

Wieland W (1982) *Platon und die Formen des Wissens*, Vandenhoeck und Ruprecht.

Wolfram S, (2023) *What Is ChatGPT Doing... and Why Does It Work?* Wolfram Media.

Xu Z, Jain S, Kankanhalli M (2024) Hallucination is inevitable: An innate limitation of large language models. *arXiv:2401.11817*.

Xie Y, Wang K, Kong Y (2021) Prevalence of research misconduct and questionable research practices: A systematic review and meta-analysis. *Sc. & Engin. ethics*, *27*(4):41.

Zhong T et al (2024) Evaluation of OpenAI o1: Opportunities and Challenges of AGI. *arXiv:2409.18486*.

Zhuo T et al (2023) Exploring AI ethics of ChatGPT: A diagnostic analysis. *arXiv*.2301.1286

Zuckert C (2009) *Plato's Philosophers: The Coherence of the Dialogues*. University of Chicago Press.

Ethical approval was not required as the study did not involve human participants

Informed consent was not required as the study did not involve human participants.

Competing Interests: the authors declare no competing interests.

Authors Contribution: the manuscript has been produced jointly by the two authors.

Data Availability: the study did not involve data sets.

And Plato met ChatGPT: an ethical reflection on the use of chatbots in scientific research writing, with a particular focus on the Social Sciences.

Reyes Calderón & Francisco Herrera

Table 1. Key concerns associated with generative AI

Author rigths and plagiarism	Cotton et al, 2023; Dehouche, 2021; Fanelli, 2015; Flanagin et al, 2023; Floridi, 2019; Gao et al, 2022; Hagendorff, 2020; Hutson, 2021; Xie et al, 2021; Pupovac & Fitzpatrick, 2018; Susnjak, 2022; Were et al, 2020.
Biases, discrimination, prejudice, racism, sexism, stereotypes, injustice, inequity, ideological biases, marginalization of minorities	Basta et al, 2019; Benthall & Haynes, 2019; Binns, 2018; Buolamwini & Gebru, 2018; Dwivedi et al, 2023; Font & Costa-Jussa, 2019; Fraser et al, 2023; Guo et al, 2023a; Hutson, 2021; Huang & Chen, 2019; Kurita et al, 2019; Maarten et al, 2020; Tschopp et al. 2023; Zhao et al, 2019; Wang et al, 2023.
Dehumanization, loss of collective human identity	Chen, 2023; Leung et al, 2021; Adham et al, 2020; Mendelson et al, 2020; Solaiman et al, 2023; Solaiman et al, 2023; Anderljung et al, 2023; Hendrycks et al, 2023.
Digital gap and technical Illiteracy	Bozkurt & Sharma, 2023; Hagendorff, 2024; Leung et al, 2021.
Hallucinations or similar phenomena	Azamfirei et al., 2023; Borji, 2023; Li, 2023; Dwivedi et al., 2023; Ji et al., 2023; Susarla et al., 2023; Shelby et al, 2023; Ray, 2023; Weidinger et al., 2022.
Harmful or inappropriate content: (offensive, pornographic, toxic or violent)	Basta et al, 2019; Carlini et al, 2020; Grassini, 2023; Gehman et al., 2020; Hendrycks et al, 2023; Illia et al, 2023; Kendall & da Silva, 2024; Kennedy et al, 2018; Mozes et al, 2023; Shelby et al, 2023; Zhuo et al., 2023.
Incomplete, inaccurate, incorrect or false information, confident justifications, fabricated references, disinformation, advertisements, deepfakes	Azaria et al, 2023; Aydin & Karaarslan, 2023; Basta et al, 2019; Bender et al, 2021; Donker, 2023; Kan et al, 2023; Megahed et al, 2023; Wang et al, 2023; Zhan et al, 2023; Zhao et al, 2019.
Maleficence	Brundage et al., 2018; D'Alessandro et al, 2023; Hagendorff, 2023; Hagendorff, 2024; Hendrycks, 2023; Mozes et al, 2023; Turchin & Denkenberger, 2020; Wang, et al, 2023; Weidinger et al, 2022; Zhan et al, 2023.
Training data quality and codification's errors	Azaria, 2022.; Azaria et al, 2023; Dwivedi et al., 2023; Su & Yang, 2023.
Mass manipulations, hate speech, social punctuation techniques, political manipulation, democratic risk	Brown, 2020; Hutson, 2021; Hartmann et al, 2023; Kennedy et al, 2018; McConnell-Ginet, 2020; Mozes et al, 2023.
Lost of work	Abdullah et al, 2022; Dwivedi et al, 2023b; Qadir et al, 2022; Lazar & Nelson, 2023

Security, privacy, data protection, unintentional extraction, leakage of confidential or private information.	Alshater, 2022; Cohen, 2013; Chen, 2023; Francke & Alexander, 2019; Grassini, 2023; Huang et al, 2022; Kobis & Mossink, 2021; Ray, 2023; Vassilev, 2023; Wang et al, 2023; Wu et al, 2023;
Sustainability	George & George, 2023; Holzapfel et al, 2022; Holzapfel et al, 2022; Gill & Kaur , 2023; Holzapfel et al, 2022; Barnett, 2023; Mannuru et al, 2023; Sastry et al. 2024; Shelby et al, 2023.
Transparency, explainability, Evaluation & Auditing, controllability, opacity and black box issues	Anderljung et al, 2023; Castelvecchi, 2016; Calderón et al, 2024; Hendrycks et al, 2023; Ji et al, 2023; Ji, 2023; Mökander et al, 2023; Wang et al, 2023.
Tendency to oligopoly and excessive power	Hao, 2020; Hagendorff, 2024; Ray, 2023; Mannuru et al, 2023; Dwivedi et al 2023; Weidinger et al, 2023.

And Plato met ChatGPT: an ethical reflection on the use of chatbots in scientific research writing, with a particular focus on the Social Sciences.

Reyes Calderón & Francisco Herrera

Table 2. Main benefits associated with generative AI for researchers

Improving efficiency: saving time and effort	
Literature review and various bibliographic sources, recommendation of articles and summaries of the state of the art	Adam, 2021; Aithal & Aithal, 2023a; Alshater, 2022; Anson & Starume, 2022; Baidoo-Anu & Ansah, 2023; Bail, 2023; Banerjee et al, 2023; Burger et al, 2023; Chen & Liu, 2023; Dergaa et al, 2023; Deroy et al, 2023; Dowling & Lucey, 2023; Else, 2023; Hill-Yardin, 2023; Kalifa & Ibrahim, 2024; Kalla, et al, 2023; Liebrenz et al, 2023; Lund et al, 2023; 2023; Ray, 2023; Salvagno et al, 2023; Sallam, 2023; Thorp, 2023.
Writing of study proposals. Structured presentation of reports and research results	
Automatic formatting, elaboration of tables, graphs and aesthetic figures Generating code to solve analytical or computational tasks	
Guidance on ethics, compliance and professional regulation	
Erasing the language barrier for non-English speakers	
Discovery of new knowledge.	
Peer-review process response; citation count	
Quality' improvement	
Obtaining expert advice	Aithal & Aithal, 2023b; Bail, 2023; Banerjee et al, 2023; Dasborough, 2023; Dwivendi et al, 2023; Gao et al, 2022; Gill & Kaur, 2023; Guo et al, 2023a; Huang & Chen, 2019; Kung et al, 2022; Lahat et al, 2023; Lund et al, 2023; Mahama et al, 2023; Pavlik, 2023; Rice et al, 2024; Ziems, 2023.
Generation of new questions based on patterns and trends and gaps in the research	
Methodological orientation, selection of methods for experimental design	
Statistical analysis, hypothesis generation, hypothesis testing, sample size, randomization, and control group allocatio	
Data visualization and exploration	
Detection and correction of errors and biases	
Cost savings	
Data collection: surveys, questionnaires or interviews	Anson & Straume, 2022; Bahrini et al, 2023; Bail, 2023; Banerjee et al, 2023; Carlini et al, 2020; Coveney et al, 2016; El badawi et al, 2024; Gefen & Arinze, 2023;

Data processing, cleaning, coding.	Grace et al, 2018; Helberger & Diakopoulos, 2023; King, 2023; McDermott et al, 2019; Rice et al, 2024; Sallam, 2023; Salvagno et al.,2023.
Interpretation of findings: conclusions and inferences	
Experience identification	
Replication of the research findings	
Collaborative platform	
Overcoming language barriers	

And Plato met ChatGPT: an ethical reflection on the use of chatbots in scientific research writing, with a particular focus on the Social Sciences.

Reyes Calderón & Francisco Herrera

Table 3. Some needed questions in relation with Hibrid Paper's acceptation

Editors must ensure that authors	
Do not include chatbots as coauthor	Bhatia & Kulkarni, 2023; Contractor et al, 2022; Flanagin et al, 2023 Gao et al, 2022; Graf & Bernardi, 2023; Gupta, 2024; Hosseini, et al, 2023; Hu, 2023; Lee, 2023; Leung et al, 2021; Liebrenz et al., 2023; Long & Magerko, 2020. Pupovac & Fanelli, 2015; Samuelson, 2023; Scerbo, 2023; Sengupta & Honavar, 2017; Stokel-Walker, 2023; Teixeira da Silva & Tsigaris, 2023; Uchendu et al, 2023; Yeo-Teh & Tang, 2023; Zhuo et al, 2023b; Zielinski et al, 2023.
Declare the use of AI generative tools in the paper, and specifications of the chatbot used: name, version, model, source.	
Take formal responsibility for all contributions made by chatbots, including accuracy and proper attribution of all cited material, and for searching for and citing omitted sources.	
Detail how the chatbot has been employed in the paper, including method of application (query structure, syntax).	
Detail specifically the contribution to the science	
Editors must ensure that referees	
Is a human expert reviewer	Baek et al., 2024; Lu et al., 2024; Si et al, 2024; Leung et al, 2021.
Have a skilled and/or interdisciplinary team able to evaluate content efficiently and accurately.	
Have appropriate tools to help them detect content generated or altered by AI, with or without it being declared.	

Have a skilled and/or interdisciplinary team able to evaluate content efficiently and accurately.	