

Thinking critically about scientific information generated by ChatGPT

Information and
Learning Sciences

Pablo Antonio Archila

School of Education, Universidad Santo Tomás, Bogota, Colombia

Brigithe Tatiana Ortiz

*Department of Chemical and Food Engineering, Universidad de los Andes,
Bogota, Colombia*

Anne-Marie Truscott de Mejía

School of Education, Universidad de los Andes, Bogota, Colombia, and

Jorge Molina

*Department of Biological Sciences, Universidad de los Andes,
Bogota, Colombia*

Received 23 April 2024
Revised 7 July 2024
8 August 2024
Accepted 8 August 2024

Abstract

Purpose – In November 2022, the commercial company, OpenAI, launched ChatGPT. Since then, university students have rapidly become regular users of this artificial intelligence (AI) platform. One reason for this is the powerful capability of this generative AI tool to produce textual content, which in many cases, is almost indistinguishable from human-generated content. Another reason is that ChatGPT easily gives anyone access to knowledge. However, there is a problem as the vast majority of its users have no idea how this AI platform works and thus overlook the importance of thinking critically about information communicated in ChatGPT. While some call for banning this generative AI tool, this study aims to provide evidence that science classrooms can become scenarios where students find explicit, concrete, and realistic opportunities to critically evaluate scientific information generated by ChatGPT.

Design/methodology/approach – An intervention study was conducted with 55 students (26 females and 29 males, 17–24 years old) during a university Spanish-English bilingual science course taught within an active learning environment. The data consist of the written critiques of the students about Spanish-English bilingual scientific texts produced by ChatGPT.

Findings – Results indicate that the intervention had a positive effect on students' abilities to construct sound arguments in Spanish and in English while judging the quality of scientific texts produced by this AI bot. Moreover, the findings suggest that the intervention enriched students' skills to make improvements to texts produced by this generative AI tool.

Originality/value – The evidence provided in this study contributes to the exploration of possibilities to help students become critical users of ChatGPT.

Keywords Argumentation, Artificial intelligence, Deep active learning, Scientific information uses

Paper type Research paper



The authors wish to thank the undergraduate students for their time and participation in this study, as well as to acknowledge financial support from the Vice-Presidency of Research and Creation, Universidad de los Andes, Bogotá, Colombia.

Information and Learning
Sciences
© Emerald Publishing Limited
2398-5348
DOI 10.1108/ILS-04-2024-0040

Introduction

In September 2022, Bill [Gates \(2023\)](#) had the opportunity to see the Chat Generative Pretrained Transformer (ChatGPT) in action. He was really impressed and described his experience as follows: “I knew I had just seen the most important advance in technology since the graphical user interface.” Two months later, [OpenAI \(2022\)](#) – the American commercial company that created this advanced generative artificial intelligence (AI) tool – introduced it as a freely available service to a mass audience. Since then, ChatGPT has rapidly received notable mediatization and gained great popularity. [Brockman \(2022\)](#), OpenAI’s president and co-founder, stressed that in only five days after its launch, one million people signed up to test it. Likewise, [Lee \(2023\)](#) emphasized that it appears to be one of the fastest online services that globe has ever adopted when compared with other services such as Instagram (2.5 months taken to reach one million users), Dropbox (8 months), Facebook (10 months), Twitter (24 months), and Netflix (41 months). Given the viral public interest in this AI platform, it was not surprising to observe that in February 2023, [OpenAI \(2023\)](#) launched ChatGPT Plus: a subscription plan available for US\$20/month.

Arguably, much of the vitality of ChatGPT is explained by its tremendous ability to pass complex written exams (e.g. United States Medical Licensing Exam [[Kung et al., 2023](#); [Mihalache et al., 2024](#)]), create art, literature, and music ([Zhu et al., 2024](#)), generate large amounts of textual content (e.g. scientific texts) ([Khalifa and Ibrahim, 2024](#); [Langin, 2024](#)), and complete language translation tasks ([Gulati et al., 2024](#); [Martin and Graulich, 2024](#)), face biometrics tasks ([DeAndres-Tame et al., 2024](#)), language editing tasks ([Alyasiri et al., 2024](#)), and software engineering tasks ([Marques et al., 2024](#); [Meyer et al., 2024](#)). Despite the potential benefits of ChatGPT, the uncritical use, misuse or abuse of this large language model raises several concerns. For instance, [Khowaja et al. \(2024\)](#) emphasize various worries, such as sustainability (a lot of energy is consumed during ChatGPT training and inferential phases), privacy (data collection and use by OpenAI are unclear), digital divide (ChatGPT is creating a digital divide among the high- and the low-income countries), and ethics (ChatGPT is not guided by ethical values).

The ethical use of ChatGPT and other generative AI tools is a major issue of concern. In fact, the use of these tools in academic and educational environments is a reiterative point of discussion in the *Second Handbook of Academic Integrity* ([Eaton, 2024](#)). Moreover, there is evidence of higher education institutions that have developed guidelines in response to generative AI ([Moorhouse et al., 2023](#)). Nevertheless, as [Zonjić \(2024\)](#) explains, “students are on the receiving end of AI-use policies, but rarely have agency over how they are developed” (p. 1). She maintains that students’ voice about the use of AI tools should be valued and made visible, while giving them opportunities to critically reflect about the strengths and weaknesses of ChatGPT. Of course, this AI tool is far from perfect. For example, in a highly cited *Science* editorial, [Thorp \(2023\)](#) says that having tested this generative AI tool with an exam and a final project of a university science course, his conclusion is that “it did well finding factual answers, but the scholarly writing still has a long way to go” (p. 313). Indeed, [OpenAI \(2022\)](#) acknowledges shortcomings relating to the performance of ChatGPT.

Also, [Thorp \(2023\)](#) notes that this AI program may invent sources/references of scientific studies that do not exist. Because not all the scientific texts generated by ChatGPT present trustworthy or accurate information, science education practices should provide students with explicit, concrete, and realistic opportunities to critically engage with ChatGPT’s texts containing scientific information. The role of science education is even more necessary if we recognize that students have rapidly become regular users of this AI platform ([Cooper, 2023](#), [Mogavi et al., 2024](#); [Tan et al., 2024](#)). It is therefore vital to foster its critical use. In this

sense, Mogavi and his colleagues (2024) argue “for a more responsible and effective integration of ChatGPT in educational contexts” (p. 18). ChatGPT seems to be a promising ally in the creation of personalized science learning experiences (Alneyadi and Wardat, 2023), as well as in the reduction of language barriers in science classrooms (Martin and Graulich, 2024), the promotion of holistic reading of socio-scientific texts (Cheung *et al.*, 2024) and the improvement of students’ mastery of scientific knowledge, motivation, and self-regulated learning (Ng *et al.*, 2024). Moreover, Gates (2023) and Lee (2023) claim that ChatGPT deserves to be used for educational purposes. However, there is insufficient research evidence to support such a claim. Much of the reason for this is that “ChatGPT research in education is still in its early stages” (Polat *et al.*, 2024, p. 77). Furthermore, in a recent systematic review, Lo *et al.* (2024) reported a lack of studies concerning the promotion of critical mindset habits that enable students to judge the quality of ChatGPT’s responses.

To address this research gap, the present study seeks to contribute to our understanding of emergent and crucial issues, namely, how to take advantage of ChatGPT in science education and how to help students become critical users of it. Specifically, the study aims to provide evidence that science classrooms can become scenarios where students find explicit opportunities to critically assess scientific information presented in this generative AI tool. To make explicit such opportunities we integrated elements of (1) deep active learning (DAL) (Matsushita, 2018), (2) scientific AI literacy, and (3) useful feedback (Winstone and Carless, 2020) throughout a one-semester-long university Spanish-English bilingual biology course. The questions addressed in this intervention study are:

- RQ1. To what extent does the university bilingual biology course give students explicit opportunities to critically evaluate scientific information presented in texts generated by ChatGPT?
- RQ2. What are students’ impressions of the intervention with regards to giving them opportunities to critically evaluate texts generated by ChatGPT?

Conceptual framing

In this section, we discuss the conceptual basis of the three mainstays (DAL, scientific AI literacy, and useful feedback) of our intervention. The first pillar relates to DAL which, according to Matsushita (2018), is aimed at promoting both content knowledge and higher-order thinking skills. Furthermore, she insists that universities ought to add the feature of depth to active learning practices, being aware that students’ acquisition of content knowledge is a prime condition for effective engagement in higher-order thinking processes. It should be noted that such acquisition will be insufficient by itself if instructors do not provide students with explicit opportunities to become engaged in higher-order thinking practices. In this sense, Mizokami (2018) claims that if university education really seeks to promote not only deep learning, but also higher-order thinking skills (e.g. argumentation), then it needs to move away from lecture-based knowledge transmission approaches since these perpetuate passive learning. In similar fashion, Nobel laureate Carl Wieman (2017) reminds us that “people do not develop true understanding of a complex subject such as science by listening passively to explanations” (p. 9). Research demonstrates that university science courses can become effective scenarios of DAL even when there are large groups (Archila *et al.*, 2022a). Nonetheless, many universities around the world still adopt instructor-centered practices, and thus, active learning seems a slogan rather than the reality of many university science courses. To give an example, in a recent study conducted in the

USA, [Idsardi et al. \(2023\)](#) examined the relationship between undergraduate science, technology, engineering, and mathematics instructors' ($N = 22$) conceptions of learning and practices. These scholars found that 12 participants primarily lectured in their courses and did not provide students with opportunities to engage in active learning.

At this point, it is valid to agree that DAL is a legitimate and desirable educational approach, but it may be asked why we decided to include it as an element of our intervention study. The reason is that we assume that the critical evaluation of scientific information presented in texts generated by ChatGPT entails both the mastery of scientific knowledge and higher-order thinking, which as we have already explained, are key features of DAL practices. This critical evaluation is fundamental because this generative AI tool can expose users to (un)intentional false and/or inaccurate scientific information ([Cappellani et al., 2024](#); [Naddaf, 2023](#); [Thorp, 2023](#)). This is why the second pillar of our intervention addresses scientific AI literacy which we define as skills to engage critically with science in AI settings. Like other types of literacies, such as scientific media literacy – which integrates scientific literacy with media literacy ([Chang Rundgren and Rundgren, 2014](#)) – and COVID-19 literacy ([Archila et al., 2021a](#)) – which combines health literacy, scientific literacy, and scientific media literacy – scientific AI literacy results from the articulation of scientific literacy with AI literacy.

[Batarseh \(2022\)](#) defines AI as a branch of computer science concerned with automating intelligence and enabling machines to perform complex tasks in complex environments. [Pretorius \(2023\)](#) and [Ng et al. \(2023\)](#) consider that the growing popularity of AI technologies in everyday life settings is one reason why AI literacy will increasingly become a major source of quality for all students. At this point, it is fair to say that “the concept of AI literacy is quite new” ([Chan and Colloton, 2024](#), p. 28). However, the construction of AI literacy frameworks is gaining global interest. [Pretorius and Cahusac de Caux's \(2024\)](#) framework is a prime example of this. They created an AI literacy framework for higher education which included five elements, namely:

- (1) Foundational – basic skills that allow for an effective use of generative AI;
- (2) Conceptual – understandings about how to locate, organize, evaluate, use, and repurpose generative AI outputs;
- (3) Social – assuming the role of interlocutor when interacting with generative AI platforms;
- (4) Ethical – acting ethically when using generative AI; and
- (5) Emotional – developing socio-emotional awareness of the use of generative AI tools.

This framework provides important insights at this time when several scholars (e.g. [Chu-Ke and Dong, 2024](#); [Kong et al., 2024](#); [Łabuz and Nehring, 2024](#)) urge education stakeholders to address the need to equip all citizens with skills to leverage advances in AI technologies from an informed and a critical point of view.

Another interesting AI literacy framework is the one proposed by [Chan and Colloton \(2024\)](#). This framework consists of five elements:

- (1) understanding AI concepts;
- (2) awareness of AI applications;
- (3) AI affectivity for human emotions;
- (4) AI safety and security; and
- (5) responsible AI usage.

Furthermore, [Chiu et al. \(2024\)](#) created a framework for K–12 AI education. This framework includes AI literacy (knowing) and AI competency (applying the knowledge), encompassing the following five aspects: technology, impact, ethics, collaboration, and self-reflection. A common feature of this framework and those of [Chan and Colloton \(2024\)](#) and [Pretorius and Cahusac de Caux \(2024\)](#) is that all of them offer a multifaceted approach. This gives us an idea of how complex the concept of AI literacy is. This type of literacy can be defined:

As a set of competencies that enables individuals to critically evaluate AI technologies; communicate and collaborate effectively with AI; and use AI as a tool online, at home, and in the workplace ([Long and Magerko, 2020](#), p. 2).

This definition is relevant in our intervention study as it includes the element of criticality that is also part of scientific literacy. To be clear, the [OECD \(2020\)](#) emphasizes that a willingness to think critically about scientific knowledge is one attribute of scientifically literate people. Additionally, this organization reiterates the need for students to become equipped to engage critically with science in AI settings.

Scientifically AI literate people are able to think critically about the scientific information presented in products (e.g. texts) created by *generative AI* – a technology that uses deep learning models to produce human-like content in response to complex and varied prompts (e.g. instructions, questions) ([Lim et al., 2023](#)). ChatGPT is in essence a generative AI technology. Academic voices are increasingly calling for the enactment of educational scenarios in which students find opportunities to critically evaluate generative AI outputs, such as texts generated by ChatGPT ([AAIN Generative AI Working Group, 2023](#); [Pretorius, 2023](#)). Recently, [Osborne and Pimentel \(2022, 2023\)](#) invited science education stakeholders to become aware that nowadays students are continuously exposed to a plethora of scientific misinformation – communication of false and/or scientific information – as one consequence of an age of misinformation. For instance, [Archila et al. \(2019, 2021b\)](#) argue that news articles can expose readers to false (e.g. confusing protozoon with virus) and/or inaccurate (e.g. use of sensationalist language) scientific information. They warn that this can lead to the public misunderstanding of science. Clearly, this is also the case of ChatGPT. To reiterate, students should be provided with explicit opportunities to critically evaluate its output.

These opportunities, we argue, can be created if students, for example, are asked both (1) to examine critically creations generated by ChatGPT and (2) to provide useful feedback on it. [Furtak et al. \(2019\)](#) note that although possibilities for engaging students in formative assessment practices are vast, summative assessment still dominates in many science classrooms around the globe. Feedback is widely recognized as the heart of formative assessment. Accordingly, useful feedback is the third pillar of our intervention. It is useful when underpinned judgments are provided ([Archila et al., 2022b](#)). This naturally implies going further than the mere use of conventional expressions (e.g. Well done!). Evidence indicates that students who give thoughtful comments and suggestions on the work of others learn constructively ([Wu and Schunn, 2023](#)). Moreover, [Topping \(2018\)](#) maintains that students demonstrate that they have constructed genuine understandings of the product under consideration (in our case texts generated by ChatGPT) when they are able to provide useful feedback.

[Winstone and Carless \(2020\)](#) observe that students cultivate their critical judgment – a prime element of critical thinking skills ([Davies and Barnett, 2015](#)) when they are prompted to provide useful feedback on the work of others (in our case ChatGPT), (1) identifying strengths, (2) weaknesses, and (3) areas for improvement. These three attributes largely explain why we included useful feedback in our intervention. [Winstone and Carless \(2020\)](#) also place emphasis on student feedback literacy – a fundamental element of assessment

literacy (Price *et al.*, 2012). These authors point out that “feedback literacy is more than just an academic skill, also being an important dimension of professional competence in the workplace” (p. 32). This undoubtedly implies that students should be provided with optimal opportunities to develop the view of feedback – a process that enables students to make sense of information and use it to enrich their learning or work strategies (Carless and Boud, 2018) – as a valuable source of lifetime learning.

Nieminen and Carless (2023) consider that “a key building block within new paradigm feedback practices is the notion of *feedback literacy*” (p. 1382, original emphasis). Student feedback literacy is defined as the group of understandings, capacities and dispositions that students need to leverage feedback processes to enrich their practice in learning or work settings (Carless and Boud, 2018). In addition, these academics claim that feedback literacy can facilitate teacher-student interactions that tend to be rare due to the expansion of mass tertiary education systems. In our study, we contend that these interactions can be fostered if the instructor uses the students’ critical judgment of texts generated by ChatGPT to engage them in informal formative assessment (also known as on-the-fly formative assessment and interactive formative assessment) – a kind of assessment for learning in which feedback is provided during the course of events (Ramnarain *et al.*, 2022). To this end, we adopted Archila *et al.*’s (2022a) notion of instructor–student argumentative interaction – a type of argumentative interaction in which instructor and students exchange and/or co-construct arguments in a symmetric atmosphere – as a platform of informal formative assessment. In this interaction, the instructor acts not as *the* authority, but as a facilitator and a challenger, taking advantage of students’ participation to interact argumentatively with them and give useful feedback. Importantly, instructor–student argumentative interaction is flexible since it can be enacted in one-on-one, small-group or whole-class discussions. We will explain later how this interaction occurred in our intervention.

Methods

Study design and setting

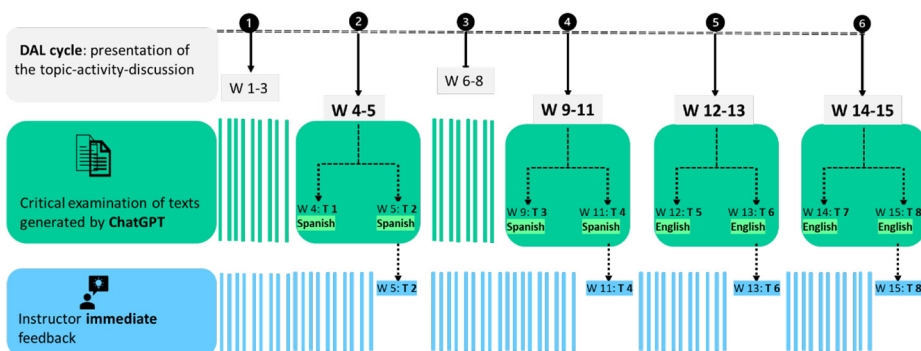
We used a quasi-experimental one-group time series design to assess the effectiveness of the educational intervention. As recommended by Cohen *et al.* (2018), we conducted repeated pre-test/post-tests throughout the intervention. They stress that this “enables the participants to become their own controls, which reduces the effects of reactivity” (p. 407). The intervention was carried out in a one-semester-long university bilingual (Spanish-English) course called the Biology of Organisms. This course was chosen because of convenient access (convenience sampling [Cohen *et al.*, 2018]), as the last author was the course instructor. This is an introductory course that is offered every semester to students in all undergraduate programs (e.g. anthropology, biology, engineering and law) at a prestigious university in Bogotá, Colombia. Following the calls to move away from Englishisation (English-only and English-always) of science to multilingual practices (Amano *et al.*, 2023; Piller *et al.*, 2022), in this course, both Spanish (the students’ first language) and English (a foreign language) are recognized as valuable educational resources for science teaching and learning (Archila *et al.*, 2022b, 2022c). Specifically, the course makes use of a preview/view/review methodology (see García, 2009, p. 214, for a fuller discussion) in a dynamic, flexible, and purposeful way through diverse translanguaging practices (Mazak and Herbas-Donoso, 2015), such as engaging students in classroom activities (e.g. the creation of phylogenetic trees) in one language and encouraging discussion in another (Archila *et al.*, 2020). It is important to emphasize here that neither a high level of English language proficiency nor a high mastery of scientific knowledge are stipulated as requirements to take this introductory course.

Participants

Our sample originally included 68 undergraduates enrolled in the Biology of Organisms course. However, 13 study participants were excluded from the analyses because they did not complete at least three out of the four pre-test/post-test series carried out during the intervention. Thus, a final sample of 55 (26 females and 29 males) participants was used. Ages ranged from 17 to 24 (M = 20, SD = 1.53). The sample of this study included participants graduated from monolingual (Spanish, *n* = 22), bilingual (Spanish-English, *n* = 24), and trilingual (e.g. Spanish-English-French, *n* = 9) schools. Ethical approval was obtained from the University’s Research Ethics Committee, approval number 1637. The study adhered to the ethical guidelines of the American Psychological Association (APA, 2017). Accordingly, written informed consent was obtained from all participants as well as the parents of the students that were below the age of consent (18 years). It is important to clarify that the non-identification of the participating students was ensured through the use of codes, for instance U36 means Undergraduate Number 36.

Intervention

The Biology of Organisms course is taught over a 16-week period and consists of two 80 lecture minutes per week. Our intervention took 15 weeks distributed in six DAL cycles (Figure 1). Specifically, as part of the intervention, the instructor implemented a DAL (Matsushita, 2018) methodology throughout the course, in which the students were given explicit in-class opportunities to become engaged in activities and discussion (small-group debates and/or whole-class debates). Following the principles of DAL, these opportunities involved both high-level complex scientific content and high-level cognitive processing (argumentative and critical thinking skills). To illustrate this idea, consider the case of the construction of phylogenetic trees – a type of visual representation of evolutionary relationships. Based on the claim that the drawing of phylogenetic trees is an even more complex process than reading or interpreting them (Young *et al.*, 2013), the instructor (1) discussed content about key features (reading, interpretation, and construction) of phylogenetic trees; (2) asked students to draw an evolutionary tree with a list of familiar organisms purposefully selected by him; and (3) created a whole-class discussion in which students produced verbal arguments, counterarguments and rebuttals (argumentative interaction) while judging the soundness of the phylogenetic trees created by their peers. He



Notes: DAL = deep active learning; W = week; T = text

Source: Figure by authors

Figure 1. Main features of the 15-week intervention

organized the verbal participation of the students to interact argumentatively with them and provide useful feedback to the whole class (informal formative assessment, Archila *et al.*, 2022a). This basic three-step (presentation of the topic-activity-discussion) cycle was used throughout the 15-week educational intervention. It is worth mentioning that, as documented in other university courses (Idsardi *et al.*, 2023), active learning practices encouraged the instructor to make the students' development of in-depth understanding (quality of learning) the priority instead of the coverage of a large variety of topics throughout the course.

In DAL cycles 2, 4, 5, and 6, the students were provided with explicit opportunities to critically judge scientific information communicated in texts generated by ChatGPT. These cycles correspond to what Pretorius and Cahusac de Caux (2024) call the "conceptual" (p. 32) element of AI literacy, as students must evaluate ChatGPT outputs. To be precise, at the beginning of each one of these four cycles the instructor told undergraduates to (1) read the text (4 in total, Appendix 1) he had previously selected which was produced by ChatGPT and was closely related to the topic that would be discussed; and (2) examine critically this text with a view to offering useful written feedback meeting three requirements:

- (1) The identification of two strengths of the text generated by ChatGPT and the production of at least two valid arguments (one per strength) – based on evidence and/or reason (Erduran *et al.*, 2022) – as well as being coherent – thus, effectively underpinning her/his judgment (Archila *et al.*, 2022b) – while answering the question, "Why do you consider this is one strength?" – the purpose of this question was to prompt students to assess critically the text;
- (2) Identification of two weaknesses (or aspects that require enhancement) and construction of at least two valid and coherent arguments while answering the question, "Why do you consider this is one weakness?"; and
- (3) Edition of the text carrying out improvements only relating to the two weaknesses identified in the second requirement.

The instructor encouraged students to revise the text as far as possible adopting critical thinking dispositions, such as open-mindedness and healthy skepticism (avoiding blind acceptance of ideas) (Davies and Barnett, 2015). The reason for this is that thinking critically about products created by generative AI is an essential feature of AI literacy (Long and Magerko, 2020). Likewise, it is reasonable to assume that students used their prior knowledge while offering useful written feedback about the text produced by ChatGPT. Hence, the instructor harnessed this to present the topic, and thus linked the text to the first step of the basic three-step (presentation of the topic-activity-discussion) cycle.

At the end of each one of the four DAL cycles, once the third step (discussion) of the basic three-step cycle was carried out, the instructor asked undergraduates to give useful written feedback about another text (4 in total) that he had previously selected which was closely related to the topic that was discussed during the cycle. Students were asked to meet the three requirements already mentioned (identification of two strengths ["Why do you consider this is one strength?"], identification of two weaknesses ["Why do you consider this is one weakness?"], and edition of the text). After this, the instructor created a whole-class discussion in which he listened to those students who voluntarily communicated their judgments of the text generated by ChatGPT. He used these judgments as a springboard to interact argumentatively with the undergraduates and provide *immediate* feedback to the whole class. Instructor feedback was intended to help students both clear up possible scientific content misunderstandings and reflect on the importance of judging products created by generative AI from a critical perspective as part of the long process of becoming

scientifically AI literate. In short, our intervention provided participants with eight explicit opportunities to critically assess scientific information presented in texts (8 in total) produced by ChatGPT. Four opportunities were given at the beginning of the DAL cycles (4 in total) and four at the end of these cycles.

Finally, the Spanish-English bilingual character of the Biology of Organisms course was another element considered in our intervention. According to [Cummins's \(2017\)](#) theory of the interdependence of literacy-related skills and knowledge across languages, providing students with purposeful opportunities to carry out activities in one language (e.g. Spanish), allows them to strengthen functional conceptual and linguistic experience that can be available in another (e.g. English). Drawing on this theory, Texts 1–4 were written in Spanish and participants were asked to offer useful written feedback in this language, while the last four (Texts 5–8) were in English and students were asked to give useful written feedback in English. To facilitate the students' gain in self-confidence, as recommended by [Archila *et al.* \(2021c\)](#), the undergraduates were provided with the opportunity to code-switch – going back and forth from one language to another in the same speech or writing ([Gardner-Chloros, 2009](#)) – between Spanish and English in their written feedback on Texts 5–8.

Research design

Data collection

Given that our study aimed at providing evidence that science classrooms can become scenarios where students are able to find concrete opportunities to critically judge scientific information presented in ChatGPT, the data corpus was composed of the undergraduates' written feedback pieces – one *piece* consisted of three elements: strengths, weaknesses, and edition – about the eight texts generated by this generative AI tool and the responses of the participating students to an anonymous online 13-item survey ([Appendix 2](#)). In total, 418 pieces were collected. As previously mentioned, during four cycles, participants were given explicit opportunities to critically assess texts generated by ChatGPT. The Microsoft Forms[®] platform was used to collect the written feedback produced by the students. At the beginning as well as at the end of each of the four cycles, participants were given access to this platform. In addition, students had access to the texts in PDF[®] files through a learning management system called: Brightspace[®]. To facilitate the data collection process, participants were asked to copy from the PDF[®] file the two parts they considered strengths of the text generated by ChatGPT and paste them in the Microsoft Forms[®] platform where they had to argue in writing twice (one per strength) while answering the question, “Why do you consider this is one strength?” This copy-paste-argue routine was repeated in the two weaknesses part. Finally, the students were required to adopt the following copy-paste-edit routine:

- copy the full text from the PDF[®] file;
- paste it in a Word[®] file; and
- edit the text generated by ChatGPT, carrying out improvements (highlighting them in yellow) only in relation to the two weaknesses they had identified.

Each student had to upload the edited file to the platform. Participants were given as much time as they needed to carry out both routines (copy-paste-argue and copy-paste-edit). During these routines, the students were asked to not seek information on the Web, or in books, or notebooks.

Each of the eight texts presented to the students was produced by ChatGPT a couple of days before the instructor carried out each of the four DAL cycles. In other words, he prompted this AI tool to write an argumentative text – a kind of text that typically follows the

macrostructure: introduction-argument-conclusion (Stavans *et al.*, 2019) – related to the topic of the cycle (e.g. ChatGPT prompt-write an argumentative text about the importance of Hardy–Weinberg equilibrium for evolution). We decided to prompt this generative AI tool to produce argumentative texts because this type of text often communicates reason-based and/or evidence-based arguments more regularly than other kinds of texts, such as, informative texts and narrative texts. Moreover, as Erduran *et al.* (2022) point out, the communication of scientific knowledge claims underpinned by the production of reason-based and/or evidence-based arguments is highly desirable.

Once ChatGPT generated the text (4 in total, two in Spanish and two in English), the instructor used the option “regenerate response” – an option that prompts this generative AI tool to produce a brand-new response – four times (16 in total), thus obtaining five different texts (20 in total) about the same topic. Then, he randomly selected two (8 in total) from these five texts. One of these texts was presented to the students at the beginning of the DAL cycle while the other was presented at the end. The length of the four texts in Spanish varied between approximately 250 and 450 words, while the four texts in English were between approximately 300 and 350 words long. To get an idea of the readability level of the eight texts presented to the students, we used the Flesch Reading Ease Score (Flesch, 1948) – a score of zero means a text is very difficult to read while a score of 100 means a text is very easy to read – for texts in English and the Fernandez-Huerta (1959) formula – “a validated Spanish equivalent [to the Flesch Reading Ease Score]” (Gorrepati and Smith, 2021, p. 142) – for texts in Spanish. These calculations were performed using the Microsoft Word[®] software (Stockmeyer, 2009) and the Legible[®] software (legible.es), respectively. The scores calculated were 52.8 for Text 1; 42.9 for Text 2; 51.4 for Text 3; 61.1 for Text 4; 15.5 for Text 5; 10.5 for Text 6; 30.5 for Text 7; and 32.1 for Text 8. This means that the texts were fairly difficult/very difficult to read.

Finally, inspired by the idea that participants’ voice is a legitimate and desirable source from which researchers can “gain feedback on an intervention” (Cohen *et al.*, 2018, p. 42), in the last cycle of the educational intervention we administered a voluntary and anonymous online 13-item survey. This instrument was in no way intended to collect precise information about participants’ satisfaction, it was merely a pragmatic tool that allowed us to receive student feedback which we believe is relevant for future improvements. Items were adapted from previous feedback surveys developed in interventions related to higher-order thinking skills (e.g. argumentation, critical thinking) (Archila *et al.*, 2021b; Brooke, 2006; Bulgren *et al.*, 2014; Mathee and Turpin, 2019) and bilingual science education (Archila *et al.*, 2021c; 2022b). The instrument was especially useful for collecting data on the respondents’:

- frequency of use of ChatGPT (Question 1 in Appendix 2);
- previous instruction in the critical evaluation of texts generated by this AI platform (Question 2 in Appendix 2);
- opinions about the intervention (Questions 3–7 in Appendix 2); and
- views about the importance of thinking critically about ChatGPT in the Biology of Organisms course (Questions 8–13 in Appendix 2).

The survey was in Spanish and undergraduates were free to respond in Spanish, English or to code-switch between these languages in answering the open-ended questions (Questions 3 and 4 in Appendix 2). Fifty-three out of the 55 undergraduates voluntarily responded.

Data analysis

Data analysis was carried out at three levels to answer our research questions. First, the bilingual character of the intervention was examined. At this level, the languages used by the

students in the written feedback pieces (418) about the judgment of texts generated by ChatGPT were examined to establish the number of feedback pieces written in Spanish, the number of pieces produced in English, and the number of pieces written using code-switching. At the second level, each piece was coded (scored) according to four criteria (codes) (Table 1). Additionally, this coding process was guided by two specifications, namely:

- (1) A correct use of scientific concepts was assumed as no errors at all in the written feedback produced by the student. For instance, “the protists organisms don’t have a true nucleus” (U27), was considered an error; and
- (2) An argument was accepted as both valid and coherent only when it was backed by reason and/or evidence (Erduran *et al.*, 2022) and it effectively supported the student’s judgment (Archila *et al.*, 2022b). Figure 2 shows an example of a written feedback piece produced by U38 and coded with the maximum score: seven.

Table 2 offers an idea of the relevance of the corpus on which this study was based as we analyzed 2,508 subpieces in total. The first author coded all the data. The second author coded written feedback pieces 1–4 and the last author coded written feedback pieces 5–8. All the statistical calculations were carried out via the *Statistical Package for the Social Sciences* (SPSS®, v28). Cohen’s kappa coefficient (Cohen, 1960) calculated was 0.94 for piece 1; 0.94 for piece 2; 0.97 for piece 3; 0.97 for piece 4; 0.93 for piece 5; 0.95 for piece 6; 0.99 for piece 7; and 0.95 for piece 8. “A coefficient of 0.75 or above [was] considered very good” (Bryman, 2016, p. 276) inter-coder agreement. All disagreements were discussed until a final consensus was achieved. Likewise, Cohen’s *d* effect size was computed to assess the magnitude of students’ gains at the end of each DAL cycle. To be specific, this computing was carried out using the written feedback piece produced by the participating students at the beginning of each cycle as the “control” condition and the piece created at the end of each cycle was treated as the “experimental” condition. To interpret our results, we followed Cohen’s (1988, pp. 25–26) small ($d = 0.2$), medium ($d = 0.5$), and large ($d = 0.8$) benchmarks.

The third level of the analysis dealt with the examination of the responses to Questions 1, 2, 5, 6, and 7 of the 13-item feedback survey (Appendix 2) using frequency counts. Furthermore, the six Likert type (1 – strongly disagree, 5 – strongly agree) items (Questions 8–13) were analyzed. Essentially, we calculated the Cronbach’s alpha coefficient which was found to be 0.88. According to George and Mallery (2022, p. 387), this alpha value is “good.” Furthermore, in the next section we comment on some of the answers to open-ended questions 3 and 4.

Findings

RQ1. To what extent does the university bilingual biology course give students explicit opportunities to critically evaluate scientific information presented in texts generated by ChatGPT?

The results of the written feedback produced by the undergraduates at the beginning of the four DAL cycles (Pre-DALC in Table 3) indicate that the participating students had trouble examining critically texts generated by ChatGPT. This is a key outcome since this demonstrates that the need to use science classrooms as scenarios where students find explicit opportunities to think critically about scientific information presented in ChatGPT is substantial. In the following paragraphs, we will now focus on the opportunities that our intervention in the Biology of Organisms course provided to the participants to judge texts generated by ChatGPT.

Table 1. Criteria used in the scoring process of the participants' written feedback pieces

Criteria	Score	Description
Use of scientific concepts	0	The student uses scientific concepts wrongly
	1 (Maximum)	The student uses scientific concepts correctly
Strengths	0	The student identifies at least one strength of the text generated by ChatGPT, but s/he does not produce at least one valid and coherent argument to underpin her/his judgment
	1	The student identifies at least one strength of the text generated by ChatGPT and constructs at least one valid and coherent argument to support her/his judgment
Weaknesses	2 (Maximum)	The student identifies at least two strengths of the text generated by ChatGPT and produces at least two valid and coherent arguments to underpin her/his judgment
	0	The student identifies at least one weakness of the text generated by ChatGPT, but s/he does not construct at least one valid and coherent argument to support her/his judgment
Enhancements	1	The student identifies at least one weakness of the text generated by ChatGPT and produces at least one valid and coherent argument to back her/his judgment
	2 (Maximum)	The student identifies at least two weaknesses of the text generated by ChatGPT and constructs at least two valid and coherent arguments to support her/his judgment
Maximum total score	0	The student edits the text generated by ChatGPT, but s/he does carry out improvements to at least one of the two weaknesses that s/he has identified earlier
	1	The student edits the text generated by ChatGPT, carrying out improvements to at least one of the two weaknesses that s/he has identified earlier
Maximum total score	2 (Maximum)	The student edits the text generated by ChatGPT, carrying out improvements to the two weaknesses that s/he has identified earlier
	7	

Source: Authors' own work

All living organisms require energy to carry out their essential functions, including growth, reproduction, and movement. Prokaryotic and eukaryotic organisms are two types of cells that differ in many ways, including how they obtain energy.

Prokaryotic organisms, such as bacteria and archaea, obtain energy through various processes, including photosynthesis, respiration, and fermentation. Photosynthetic prokaryotes use sunlight to synthesize organic molecules, such as glucose, from carbon dioxide and water. Non-photosynthetic prokaryotes obtain energy by breaking down organic molecules, such as glucose, through respiration or fermentation. Respiration is a more efficient process that uses oxygen to produce more ATP (adenosine triphosphate), the molecule that cells use as their primary energy source. Fermentation, on the other hand, is a less efficient process that does not require oxygen and produces less ATP.

Eukaryotic organisms, including animals, plants, fungi, and protists, obtain energy through a variety of processes, including photosynthesis, respiration, and fermentation. Photosynthetic eukaryotes, such as plants and algae, use sunlight to synthesize organic molecules from carbon dioxide and water. Non-photosynthetic eukaryotes obtain energy by breaking down organic molecules, such as glucose, through respiration or fermentation. Respiration is a more efficient process that uses oxygen to produce more ATP, while fermentation is a less efficient process that produces less ATP.

However, eukaryotic organisms have a unique feature that allows them to obtain energy in a different way than prokaryotic organisms. Eukaryotes have specialized organelles called mitochondria that are responsible for producing ATP through respiration. Mitochondria are believed to have evolved from an ancient prokaryotic organism that was engulfed by a eukaryotic cell. The symbiotic relationship between the two cells allowed the eukaryotic cell to obtain energy more efficiently, leading to the evolution of more complex eukaryotic organisms.

In conclusion, both prokaryotic and eukaryotic organisms obtain energy through various processes, including photosynthesis, respiration, and fermentation. However, eukaryotic organisms have a unique advantage due to the presence of mitochondria, which allows them to produce more ATP through respiration. The evolution of mitochondria in eukaryotic cells is a testament to the adaptability of living organisms and their ability to develop specialized structures to better survive and thrive in their environment.

I think this is good because it explains that the photosynthesis processes are anabolic processes, as it says that this synthesizes organic molecules from some precursors

I would change just respiration, for aerobic cellular respiration, and add after fermentation, anaerobic cellular respiration. This because it shows the concepts as if they were totally different, but when there is no oxygen, anaerobic cellular respiration is done, which results in fermentation

[...] Aerobic cellular respiration is a more [...] Fermentation, which is anaerobic cellular respiration, on the other hand, [...]

I think this explanation could be improved because as I understand the endosymbiont theory, an archaea, not a eukaryotic cell, engulfed a bacteria, which later evolved into mitochondria. The result of this process was the eukaryotic cell

[...] engulfed by archaea. The endosymbiotic relationship between them then resulted in the eukaryotic cell [...]

I think this is good because it explains that the symbiotic relationship between the archaea and the bacteria that later became the mitochondria, is a very important evidence of evolution

Notes: Strengths in blue; weaknesses in red; enhancements in green

Source: Figure by authors

Figure 2. Written feedback given by U45 about a text generated by ChatGPT (Text 6 in Appendix 1)

Table 2. Details of the 2,508 subpieces of the 418 pieces coded

Subpiece	WFP 1 (n = 54)	WFP 2 (n = 54)	WFP 3 (n = 49)	WFP 4 (n = 49)	WFP 5 (n = 53)	WFP 6 (n = 53)	WFP 7 (n = 53)	WFP 8 (n = 53)	Total (N = 418)
S	108	108	98	98	106	106	106	106	836
W	108	108	98	98	106	106	106	106	826
E	108	108	98	98	106	106	106	106	836
Total	324	324	294	294	318	318	318	318	2,508

Notes: WFP = written feedback piece; S = strengths; W = weaknesses; E = enhancements

Source: Authors' own work

In general, our results suggest that the written feedback produced by the students at the end of each one of the four DAL cycles (Post-DALC in Table 3) was slightly better than that produced at the beginning. To be specific, the intervention showed a small/medium effect size. These are in essence positive outcomes for several reasons. First, undergraduates were presented with a highly complex task. One fundamental condition for the safe use of ChatGPT is that users have to “have expert[i]se to verify that the output is accurate” (UNESCO, 2023, p. 6). The noun (1) “expertise,” the verb (2) “verify,” and the adjective (3) “accurate” are defined in the Oxford English Dictionary (2024) as (1) “expert knowledge or skill in a particular subject, activity or job,” (2) “to check that something is true or accurate,” and (3) “correct and true in every detail,” respectively. These three definitions give us an idea of how complex the task of critically judging ChatGPT-generated texts can be. Expertise

Table 3. Means, standard deviations and total scores of the written feedback pieces produced per cycle

WFP	<i>n</i>	USC Mean (SD)	S Mean (SD)	W Mean (SD)	E Mean (SD)	Total Mean (SD)	<i>d</i>
<i>WFP in Spanish</i>							
1 (Pre-DALC 1)	54	0.53 (0.50)	1.31 (0.82)	1.29 (0.71)	0.38 (0.62)	3.54 (1.92)	
2 (Post-DALC 1)	54	0.94 (0.23)	1.37 (0.65)	1.46 (0.66)	0.59 (0.68)	4.37 (1.29)	0.50
3 (Pre-DALC 2)	49	0.65 (0.48)	0.77 (0.82)	0.81 (0.75)	0.32 (0.51)	2.57 (1.80)	
4 (Post-DALC 2)	49	0.75 (0.43)	1.04 (0.76)	1.32 (0.74)	1.65 (0.69)	4.77 (2.01)	1.22
<i>WFP in English</i>							
5 (Pre-DALC 3)	53	0.52 (0.50)	1.50 (0.74)	1.30 (0.78)	0.69 (0.72)	4.01 (1.80)	
6 (Post-DALC 3)	53	0.56 (0.50)	1.62 (0.68)	1.37 (0.81)	0.88 (0.84)	4.45 (1.85)	0.24
7 (Pre-DALC 4)	53	0.39 (0.49)	1.20 (0.84)	0.92 (0.85)	0.52 (0.74)	3.05 (2.21)	
8 (Post-DALC 4)	53	0.52 (0.50)	1.52 (0.69)	1.18 (0.73)	0.79 (0.76)	4.03 (1.88)	0.47

Notes: WFP = written feedback piece; DALC = deep active learning cycle; USC = use of scientific concepts; S = strengths; W = weaknesses; E = enhancements; *d* = effect size

Source: Authors' own work

seems to play a determinant role. Therefore, it is reasonable to consider that the small/medium effect size reported here is a positive result since our intervention was carried out in an introductory (not an advanced) biology course in which students are not necessarily expected to develop “expert” biological knowledge. By the same token, the concrete action of providing students with explicit opportunities to critically judge scientific information presented in texts generated by ChatGPT could be seen as a step in the right direction, toward the exploration of genuine possibilities to help students assess scientific information communicated through generative AI technologies from a critical point of view.

Although this article does not aim to document false and/or inaccurate scientific information presented in ChatGPT, [Table 4](#) shows examples of these weaknesses that the instructor discussed with the students (whole-class discussion) at the end of each Post-DALC. We included this table to briefly illustrate how ChatGPT can foster the public misunderstanding of science. Clearly, the fact that we have asked students to provide arguments to support their judgment about strengths and weaknesses in texts generated by this AI tool, increased the complexity of the task because they had to produce valid and coherent arguments to support their judgment. Also, the participants were asked to edit the texts, carrying out improvements.

Having demonstrated that our intervention contributed to the Biology of Organisms course in providing undergraduates with explicit, concrete, and realistic opportunities to critically judge scientific information presented in ChatGPT, it is important to note that our findings suggest that this intervention also gave undergraduates opportunities to use their Spanish-English bilingual repertoires ([Figure 3](#)). Specifically, guided by the theory of the interdependence of literacy-related skills and knowledge across languages ([Cummins, 2017](#)), in our intervention, participants were asked to offer written feedback in Spanish for Texts 1–4 and in English (or code-switch) for Texts 5–8:

RQ2. What are students' impressions of the intervention with regards to giving them opportunities to critically evaluate texts generated by ChatGPT?

We believe that hearing participants' impressions about being involved in the activities of critiquing ChatGPT-generated texts is beneficial not only to have an idea of the contribution

Table 4. Examples of false or inaccurate scientific information presented in Texts 2, 4, 6 and 8

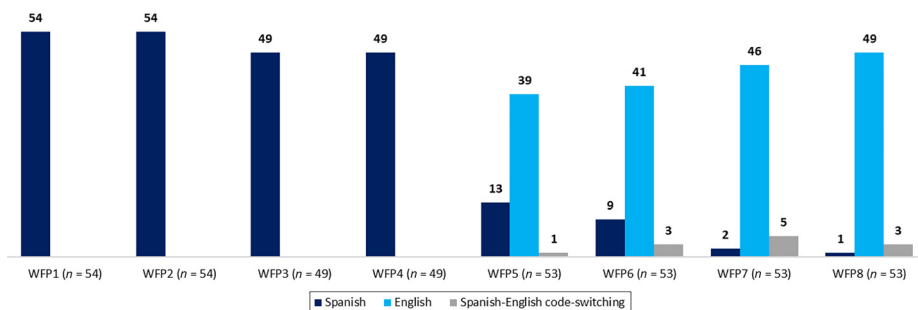
Text Type	Example	Commentary
2	Inaccurate "Hardy-Weinberg equilibrium is a key concept in genetics" (Text 2 in Appendix 1)	In this example, ChatGPT produced inaccurate information. A more accurate view was developed by Mayo (2008) who claims that "HWE is the fundamental starting point for all <i>population-genetical investigation</i> [emphasis added], whether the goal is detection or estimation of the effects of all the forces that disrupt HWE" (p. 254)
False	"Hardy-Weinberg equilibrium [...] describes the distribution of gene frequencies in a population" (Text 2 in Appendix 1)	This example shows that ChatGPT generated false information. Mayo (2008) reminds us that "Hardy-Weinberg equilibrium (HWE) is the state of the genotypic frequency of two <i>alleles</i> [emphasis added] of one gene locus" (p. 249)
4	Inaccurate "In summary, the dikaryotic phase is essential for sexual reproduction in organisms that exhibit it, such as plants, animals, protists and fungi" (Text 4 in Appendix 1)	This is inaccurate because only Ascomycota and Basidiomycota fungi have a clear and long dikaryotic phase. In this sense, Stajich et al. (2009) define Dikarya as "a subkingdom embracing the two largest fungal phyla, Ascomycota and Basidiomycota, home to ~98% of described fungi" (p. 842)
False	"Fungi reproduce asexually through the production of haploid spores, but sexual reproduction requires the fusion of two haploid nuclei to produce the dikaryotic phase, which in turn will produce the diploid spores" (Text 4 in Appendix 1)	This ChatGPT output presents false information. Nieuwenhuis et al. (2013) explain that fungi after sexual reproduction produce a dikaryotic phase, then a diploid phase followed by meiosis to produce "haploid" (p. 446) spores
6	Inaccurate "The symbiotic relationship between the two cells allowed the eukaryotic cell to obtain energy more efficiently, leading to the evolution of more complex eukaryotic organisms" (Text 6 in Appendix 1)	This can be judged as inaccurate if we recognize that prokaryotic organisms like bacteria and archaea are also complex organisms (Stevens and Arkin, 2000)
False	"Non-photosynthetic prokaryotes obtain energy by breaking down organic molecules, such as glucose, through respiration or fermentation [...] Non-photosynthetic eukaryotes obtain energy by breaking down organic molecules, such as glucose, through respiration or fermentation" (Text 6 in Appendix 1)	This is false. According to Millar et al. (2011) and Soo et al. (2017) photosynthetic prokaryotes and eukaryotes also obtain energy through respiration or fermentation

(continued)

Table 4. Continued

Text Type	Example	Commentary
8	Inaccurate “Bacteria and Archaea are both prokaryotes, meaning that they lack a membrane-bound nucleus and other complex cellular structures found in eukaryotes” (Text 8 in Appendix 1)	This ChatGPT output is inaccurate since “in the past 10 years, numerous studies have demonstrated that bacteria compartmentalize many cellular processes into subcellular structures, including organelles” (Greening and Lithgow, 2020, p. 677)
False	“[...] domain Archaea [...] split into two separate domains” (Text 8 in Appendix 1) as an example of phylogenetic relationships different to the three-domain system	This example serves to illustrate that the ChatGPT-generated text contains false information. To be clear, Raymann et al. (2015) provide evidence for the claim that Archaea separated into two domains is the explanation of the three-domain system. Only when Archaea and Eukaryotes are not split, both together are explaining the two-domain system

Source: Authors’ own work



Note: WFP = written feedback piece

Source: Figure by authors

Figure 3. Number of undergraduates who produced their written feedback pieces in Spanish, in English or in Spanish-English code-switching

of our intervention study, but for future improvements. The results of the 13-item survey reveal that nearly all the respondents of the survey (52/53) considered that the introductory biology course provided them with opportunities to critically assess ChatGPT-generated texts (Q3 in [Appendix 2](#)). Students reaffirmed this with comments, such as “The course gave me various examples of how ChatGPT can disseminate information that is not entirely correct,” “The course showed me that ChatGPT can generate false information,” “In this course I had the opportunity to examine texts generated by ChatGPT before and after discussing the topics,” and “This course gave me the opportunity to realize that there were errors in the texts. I used to hear that ChatGPT was never wrong.” Nevertheless, if these comments are not sufficient to evidence the valuable contribution of our intervention study, it is important to bear in mind that a large fraction of the respondents (47/53) commented that they never (35/53) or infrequently (12/53) had the opportunity to explicitly reflect on the importance of using this generative AI tool critically in other university courses (Q7 in [Appendix 2](#)). This contribution is even more relevant if we take into consideration that only six out of the 53 undergraduates who answered the survey had received instruction in the critical examination of texts generated by ChatGPT (Q2 in [Appendix 2](#)).

Additionally, 35 out of the 53 respondents considered that the copy-paste-edit routine about weaknesses of the texts generated by ChatGPT, was the most difficult to develop for them (Q4 in [Appendix 2](#)), followed by the copy-paste-argue routine about weaknesses of the texts (26/53) and the copy-paste-argue routine about strengths of the texts (15/53). Some comments include: “At first glance it seemed to me that the texts were completely fine. So, I had to review them thoroughly in order to identify weaknesses,” “Once I was able to identify weaknesses, I did not always know how to improve them,” and “It is complex to improve something that apparently seems correct.” Moreover, an interesting finding of the 13-item survey is that the top two difficulties respondents came up against in producing written feedback pieces in Spanish [mastery of scientific knowledge (35/53) > scientific vocabulary (20/53)] (Q5 in [Appendix 2](#)) were the same as in English [mastery of scientific knowledge (30/53) > scientific vocabulary (27/53)] (Q6 in [Appendix 2](#)).

Finally, [Table 5](#) summarizes the survey of respondents’ average scores (maximum possible = 5) along with the standard deviations on items from 8 to 13. The respondents’ average scores varied between 4.07 and 4.60 with a mean of 4.43. This value corresponds to

Table 5. Descriptive statistics of survey questions 8–13

#	Question	Mean	S.D.	Cronbach's alpha if item deleted
8.	The activities of critiquing texts generated by ChatGPT helped me become aware of the importance of using this generative AI tool critically	4.54	0.82	0.84
9.	The activities of critiquing texts generated by ChatGPT helped me realize that this generative AI tool can expose people to false and/or inaccurate scientific information	4.60	0.79	0.83
10.	The texts generated by ChatGPT and examined in the activities were related to topics discussed in the Biology of Organisms course	4.49	0.93	0.85
11.	The activities of critiquing texts generated by ChatGPT required me to use my prior knowledge	4.45	0.84	0.85
12.	The activities of critiquing texts generated by ChatGPT required me to use what I have learned in the Biology of Organisms course	4.47	0.89	0.84
13.	The critical revision of texts generated by ChatGPT requires the adoption of healthy skepticism	4.07	1.05	0.93

Source: Authors' own work

the “agree” (Bakar *et al.*, 2022, p. 7) choice. Hence, it is appropriate to claim that our intervention is a promising possibility to help science classrooms become scenarios where students find opportunities to critically judge scientific information presented in ChatGPT. Furthermore, it is worth noting that students scored the item “The activities of critiquing texts generated by ChatGPT helped me realize that this chatbot can expose people to false and/or inaccurate scientific information” (Item 9 in Table 5) the most (4.60). This result is particularly encouraging if we accept the fact that the majority of the undergraduates (41/53) said they were users of ChatGPT with greater or lesser frequency (Q1 in Appendix 2).

Discussion and educational implications

The university bilingual biology course gave students explicit opportunities to critically judge scientific information presented in texts produced by ChatGPT (RQ1)

In answer to our RQ1, the findings suggest that the biology course effectively gave undergraduates from different disciplines opportunities to evaluate scientific information presented in texts generated by ChatGPT from a critical point of view. In total, the participating students evaluated eight argumentative texts. Pretorius and Cahusac de Caux (2024) remind us that the evaluation of generative AI outputs is an element of AI literacy. In addition, the results indicate that the DAL cycles helped students to improve their skills to critically assess ChatGPT-generated texts (Table 3). However, this improvement was insufficient since they continued to have trouble carrying out copy-paste-argue routines about strengths and weaknesses of the texts as well as copy-paste-edit routines. This corroborates the claim that *expertise* is a vital condition to being able to critically examine the quality of ChatGPT output (UNESCO, 2023). Of course, it is unlikely that an introductory science course prepares *experts*. It is, therefore, plausible to say that our outcomes add empirical evidence to Archila *et al.* (2024a) and Osborne and Pimentel's (2022, 2023) reflection that given the fact that the level of expertise that students can develop

in any science course is limited, science education should focus its efforts on the preparation of students to engage critically with digital scientific information.

One important educational implication that can be drawn from our results is that the promotion of scientific AI literacy should become a high priority for science education stakeholders. Scientifically AI literate people – citizens who are able to engage critically with science in AI settings – will become increasingly in demand. In the case of university students, [Chan and Colloton \(2024, p. 248\)](#) “predict that employers will seek candidates who are AI literate.” Also, they “predict that the primary users of AI will be the more educated members of society, regardless of their economic backgrounds, with a significant proportion being younger individuals” (p. 241). In view of these predictions, we consider that our study is relevant, timely, and responsive. Likewise, our findings shed light on the promotion of scientific AI literacy and reaffirm the necessity of developing scientific AI literacy frameworks that leverage existing AI literacy frameworks ([Chan and Colloton, 2024](#); [Chiu et al., 2024](#); [Pretorius and Cahusac de Caux, 2024](#)). Every evidence-based action to encourage critical engagement with science in AI environments is essentially a welcomed step forward.

ChatGPT can expose people to false and/or inaccurate scientific information. This is one of the multiple reasons why [Thorp \(2023\)](#) invites us to view this generative AI tool with caution and not to overestimate its capabilities. The eight ChatGPT-generated argumentative texts ([Appendix 1](#)) are instructive examples of this since upon first glance they appeared to be sound. These met with various of the elements that characterize this type of texts, such as (1) paragraphs-demarcated macrostructure (introduction-argument-conclusion [[Stavans et al., 2019](#)]) and (2) logical use of connectors ([Archila, 2015](#); [Archila et al., 2017](#)). Nonetheless, portions of false and/or inaccurate scientific information become evident once one carries out a detailed examination ([Table 4](#)). This clearly explains why [UNESCO \(2023\)](#) strongly recommends that users of ChatGPT should be aware that it “cannot distinguish between right and wrong, true and false” (p. 11). Thus, our study is unique in showing that a science course can be a scenario in which students find purposeful opportunities to become aware that although ChatGPT is a powerful tool capable of generating science-related argumentative texts almost indistinguishable from human-generated texts, the adoption of healthy skepticism of the scientific information presented in these texts is crucial.

Healthy skepticism and open-mindedness are two critical thinking dispositions among many. Our results ([Table 3](#)) indicate that students were given opportunities to examine both the strengths (open-mindedness) and the weaknesses (healthy skepticism) of the texts generated by ChatGPT. Thus, students assumed the (fictional) role of assessors of ChatGPT-generated texts by producing written feedback about these texts. This is not only in line with the view that useful feedback practices can be implemented in versatile ways ([Winstone and Carless, 2020](#); [Wu and Schunn, 2023](#)), but also provides insights into the exploration of possibilities about how to use ChatGPT in educational settings. Specifically, the fact that undergraduate students were given opportunities to play the role of assessors of science-related texts written by ChatGPT expands the spectrum of the global discussion, which mainly focuses on the use of this generative AI tool to complement instructor assessment of undergraduates’ academic writing ([Lu et al., 2024](#)).

Finally, the results ([Figure 3](#)) demonstrate the bilingual character of our intervention. This is an evidence-based response to the invitation of [Amano et al. \(2023\)](#) and [Piller et al. \(2022\)](#) to move from the Englishisation of science to multilingual practices. In theory, the fact that ChatGPT can communicate with users in multiple languages opens a great possibility for the promotion of multilingual scientific practices. In practice, nonetheless, this would be unproductive if the aim of strengthening global knowledge production continues to be guided by

monolingual ideologies (e.g. English as a lingua franca). Likewise, the outcomes appear to support Cummins's (2017) view of languages as a connected entity that can be exploited if the opportunities students find are not limited to communication in one language but they are able to use their bilingual repertoires. In our intervention, students were presented with texts generated by ChatGPT in Spanish and in English. Also, they were asked to produce written feedback in these languages. In this regard, the results contribute to our understanding of how ChatGPT can be used as a generative AI tool in creating optimal opportunities for students to actively participate in multilingual science education scenarios (Martin and Graulich, 2024). It is worth adding here that many researchers (e.g. Archila *et al.*, 2024b; Charamba, 2022; Lemmi and Pérez, 2024) recommend that instructors should move away from monolingual toward multilingual practices that encourage students to use, value, and cultivate their multilingual capital. One implication is that the potential of ChatGPT to generate multilingual outputs needs to be assumed by science education stakeholders as a possibility to make multilingual scientific practices a reality rather than mere rhetoric.

Participants' impressions of the intervention (RQ2)

The results (Table 5) seem to indicate that participants had positive impressions of the intervention. The importance of these outcomes lies in the fact that they reflect the voice of the students. It is especially encouraging that the undergraduates appeared to recognize that the intervention helped them develop a more informed and critical view of the use of ChatGPT. One educational implication of this outcome might be to turn our gaze toward Zonjić's (2024) reasoning that education initiatives related to AI literacy should value the voice of the student, as we have much to learn from their experiences with generative AI platforms; after all, "students are the key users of AI chatbots in university settings" (p. 1). As such, this means that the process of constructing and implementing AI literacy frameworks (Chan and Colloton, 2024; Chiu *et al.*, 2024; Pretorius and Cahusac de Caux, 2024) and comprehensive educational guidelines pertaining to the use of generative AI tools (Moorhouse *et al.*, 2023) would benefit from these experiences.

The inclusion of suggestions for how to revise is another relevant feature of useful feedback. In our case, students were asked to edit the ChatGPT-generated texts carrying out improvements only to the weaknesses they had identified earlier. In other words, students were given opportunities to (1) identify weaknesses; (2) argue why they considered these were weaknesses; and (3) correct the errors of ChatGPT-generated texts. The outcomes of the intervention as well as the comments of the students in the anonymous survey seem to suggest that this was a complex and challenging task. Importantly, students appeared to agree that this type of task required them to use their prior knowledge and what they had learned in the Biology of Organisms course. Hence, these findings add evidence to the recent claim that involving students in the edition of texts generated by ChatGPT could be an option for the use of this generative AI tool for educational purposes. Cooper (2023), for instance, comments that in his classes students are instructed to not reference ChatGPT verbatim in assessments. "Students may first generate essay text in ChatGPT and subsequently insert key references mentioned in class" (p. 450). Arguably, the use of ChatGPT in this way is unlikely to be formative for students because, as he observes, "it seems more like an essay hack" (p. 450). This example gives us an idea of the educational value of our intervention since participants were provided with opportunities to think critically about scientific information presented in this AI tool.

It may be obvious to point out that judging the quality of scientific texts produced by ChatGPT critically is crucial for users to fully exploit the generative potential of this AI tool. Nevertheless, as Lo *et al.* (2024) note, to date there is little empirical evidence available

about how to prepare students to critically use this AI tool. We argue that providing students with explicit, concrete, and realistic opportunities to critically assess scientific information presented in ChatGPT is a legitimate and desirable first step that can be taken in science classrooms. Naturally, it is hard to imagine that students could find such opportunities in instructor-centered classrooms. The problem is that, as [Idsardi et al. \(2023\)](#) observe, active learning practices are virtually absent from many classrooms. For this reason, it makes sense to ensure that active learning is a prime condition of any educational practice aimed at the preparation of critical users of generative AI technologies. Hence, the findings of our study contribute to the research literature by reporting an educational intervention that integrates principles of DAL, scientific AI literacy, and useful feedback.

Conclusions

Two overarching conclusions may be drawn from this intervention study. The first is that our intervention, that combined elements of DAL, scientific AI literacy, and useful feedback, is a valuable contribution to the exploration of ways to provide students with concrete and realistic opportunities to critically evaluate the quality of scientific information presented in texts written by ChatGPT. This is paramount because ChatGPT-produced texts can contain false and/or inaccurate scientific information ([Cappellani et al., 2024](#); [Naddaf, 2023](#); [Thorp, 2023](#)). The second conclusion is that students tend to be willing to cultivate their scientific AI literacy skills when given the opportunity to do so. We hope this conclusion will inspire the creation of student-centered scientific AI literacy practices.

Limitations and scope for future research

Although, our study provided a timely and rare opportunity for students to critically examine science-related texts generated by ChatGPT, this is not without limitations. First, the intervention was carried out in only one introductory biology course. Thus, we acknowledge that the inclusion of other scientific disciplines as well as other course profiles (e.g. advanced science courses) would provide a broader set of evidence. A second limitation is that we did not collect data related to participants' prior experience with generative AI platforms and English language proficiency levels, which may be contextual factors that influence our results. While this has never been the intention of this study, we recognize that an in-depth analysis of these factors would help determine which, if any, influence the evaluation of science-related texts written by ChatGPT. Another shortcoming is that participants were students from diverse education majors (e.g. Biology, Industrial Engineering). This clearly limits the generalizability of the findings. These should therefore be taken more as a point of departure for future studies interested in the use of generative AI tools in science education settings and less as illustrative of specialized science courses. Additionally, an inherent limitation is that we focused on argumentative texts about topics (e.g. the Hardy–Weinberg equilibrium) that are probably less familiar to the students than others, such as climate change and COVID-19. Of course, it would be interesting to explore this intervention using more familiar topics to investigate the influence of familiarity on students' ability to critically assess ChatGPT-generated texts.

Generative AI never ceases to amaze us. Recently, the powerful capabilities of ChatGPT have revived the debate about the use of AI tools. One example of this is the case of the *Nature* and Springer Nature journals (2023) which ask authors using large language models to document this use in a section of the manuscript. We believe that science education has a vital role to play in fostering scientific AI literacy. Further work about how to promote this type of literacy is clearly needed and is likely to offer insights about how best to prepare students to engage critically with science in AI settings and to become aware of the

importance of reflecting about possible ethical issues (e.g. copyright infringement). As always, the use of science-related generative AI tools will continue to raise questions. Therefore, a future line of research could be to explore educational strategies to cultivate both open-mindedness and attitudes of healthy skepticism that help students to acknowledge both the strengths and the weaknesses of generative AI tools. There is much work to be done, after all, “the age of AI has begun” (Gates, 2023, p. 1).

References

- AAIN Generative AI Working Group (2023), “AAIN generative artificial intelligence guidelines”, Australian Academic Integrity Network, doi: [10.26187/sbwr-kq49](https://doi.org/10.26187/sbwr-kq49).
- Alneyadi, S. and Wardat, Y. (2023), “ChatGPT: revolutionizing student achievement in the electronic magnetism unit for eleventh-grade students in emirates schools”, *Contemporary Educational Technology*, Vol. 15 No. 4, pp. 1-18.
- Alyasiri, O.M., Salman, A.M., Akhtom, D. and Salisu, S. (2024), “ChatGPT revisited: using ChatGPT-4 for finding references and editing language in medical scientific articles”, *Journal of Stomatology, Oral and Maxillofacial Surgery*, doi: [10.1016/j.jormas.2024.101842](https://doi.org/10.1016/j.jormas.2024.101842).
- Amano, T., Berdejo-Espinola, V., Akasaka, M., de Andrade Junior, M.A., Blaise, N., Checco, J., Çilingir, F.G., Citegetse, G., Corella Tor, M., Drobnik, S.M. and Giakoumi, S. (2023), “The role of non-English-language science in informing national biodiversity assessments”, *Nature Sustainability*, Vol. 6 No. 7, pp. 845-854.
- American Psychological Association (2017), “Ethical principles of psychologists and code of conduct (2002, amended effective June 1, 2010, and January 1, 2017)”, available at: www.apa.org/ethics/code/index.aspx
- Archila, P.A. (2015), “Uso de conectores y vocabulario espontaneo en la argumentación escrita: aportes a la alfabetización científica”, *Revista Eureka Sobre Enseñanza y Divulgación de Las Ciencias*, Vol. 12 No. 3, pp. 402-418.
- Archila, P.A., Ortiz, B.T. and Truscott de Mejía, A.-M. (2024a), “Beyond the passive absorption of information: engaging students in the critical reading of scientific articles”, *Science and Education*, doi: [10.1007/s11191-024-00507-1](https://doi.org/10.1007/s11191-024-00507-1).
- Archila, P.A., Luna-Calderón, P. and Mesa-Piñeros, M. (2017), “El empleo espontáneo de conectores y vocabulario relacionado con las ciencias”, *Revista Eureka Sobre Enseñanza y Divulgación de Las Ciencias*, Vol. 14 No. 1, pp. 3-23.
- Archila, P.A., Molina, J. and Truscott de Mejía, A.-M. (2019), “Promoting undergraduates’ awareness of the importance of thinking critically about false or inaccurate scientific information presented in news articles”, *Revista Eureka Sobre Enseñanza y Divulgación de Las Ciencias*, Vol. 16 No. 3, pp. 1-27.
- Archila, P.A., Molina, J. and Truscott de Mejía, A.-M. (2020), “Introducing undergraduates to the nature of science through the co-construction of evolutionary trees”, *Research in Science Education*, Vol. 50 No. 5, pp. 1917-1942.
- Archila, P.A., Restrepo, S., Truscott de Mejía, A.-M., Rueda-Esteban, R. and Bloch, N.I. (2022a), “Fostering instructor-student argumentative interaction in online lecturing to large groups”, *Revista Eureka Sobre Enseñanza y Divulgación de Las Ciencias*, Vol. 19 No. 1, pp. 1-16.
- Archila, P.A., Barbosa, V., Gravier, G., Levy, L., Ortiz, B.T., Wilches, L., de Mejía, A.M.T. and Restrepo, S. (2022b), “Integrating peer feedback and instructor feedback to support the construction of bilingual scientific argument maps”, *International Journal of Science Education*, Vol. 44 No. 15, pp. 2283-2305.
- Archila, P.A., Danies, G., Calderón, F., Coronado, I., Truscott de Mejía, A.-M. and Restrepo, S. (2024b), “Student perceptions of bilingual teaching practices: evidence from a Colombian university”, *Journal of Language, Identity and Education*, doi: [10.1080/15348458.2024.2345297](https://doi.org/10.1080/15348458.2024.2345297).

- Archila, P.A., Danies, G., Molina, J., Truscott de Mejía, A.-M. and Restrepo, S. (2021a), "Towards covid-19 literacy", *Science and Education*, Vol. 30 No. 4, pp. 785-808.
- Archila, P.A., Danies, G., Truscott de Mejía, A.-M., Restrepo, S. and Molina, J. (2022c), "Students attending monolingual, bilingual, or trilingual schools at secondary level: does it influence their academic performance in a university bilingual biology module?", *International Journal of Bilingual Education and Bilingualism*, Vol. 25 No. 8, pp. 3098-3111.
- Archila, P.A., Molina, J., Danies, G., Truscott de Mejía, A.-M. and Restrepo, S. (2021b), "Providing undergraduates with opportunities to explicitly reflect on how news articles promote the public (mis)understanding of science", *Science and Education*, Vol. 30 No. 2, pp. 267-291.
- Archila, P.A., Molina, J. and Truscott de Mejía, A.-M. (2021c), "Fostering bilingual scientific writing through a systematic and purposeful code-switching pedagogical strategy", *International Journal of Bilingual Education and Bilingualism*, Vol. 24 No. 6, pp. 785-803.
- Bakar, A., Adams, F.H., Tetteh-Osei, F., Ochour, B.K., Ansah, P.O. and Asabere, I.K. (2022), "Teachers' technological, pedagogical and content knowledge in the junior high school social studies curriculum", *Universal Journal of Social Sciences and Humanities*, Vol. 2 No. 1, pp. 34-45.
- Batarseh, F.A. (2022), "Artificial intelligence", in Schintler, L.A. and McNeely, C.L. (Eds), *Encyclopedia of Big Data*, Springer, Cham, pp. 45-48.
- Brockman, G. (2022), "ChatGPT just crossed 1 million users; it's been 5 days since launch [tweet]", Twitter, available at: <https://twitter.com/gdb/status/1599683104142430208>
- Brooke, S.L. (2006), "Using the case method to teach online classes", *International Journal of Teaching and Learning in Higher Education*, Vol. 18 No. 2, pp. 142-149.
- Bryman, A. (2016), *Social Research Methods*, 5th ed., Oxford University Press, Oxford.
- Bulgren, J.A., Ellis, J.D. and Marquis, J.G. (2014), "The use and effectiveness of an argumentation and evaluation intervention in science classes", *Journal of Science Education and Technology*, Vol. 23 No. 1, pp. 82-97.
- Cappellani, F., Card, K.R., Shields, C.L., Pulido, J.S. and Haller, J.A. (2024), "Reliability and accuracy of artificial intelligence ChatGPT in providing information on ophthalmic diseases and management to patients", *Eye*, Vol. 38 No. 7, pp. 1368-1373.
- Carless, D. and Boud, D. (2018), "The development of student feedback literacy: enabling uptake of feedback", *Assessment and Evaluation in Higher Education*, Vol. 43 No. 8, pp. 1315-1325.
- Chan, C.K.Y. and Colloton, T. (2024), *Generative AI in Higher Education: The ChatGPT Effect*, Routledge, Abingdon.
- Chang Rundgren, S.-N. and Rundgren, C.-J. (2014), "SSI pedagogic discourse", in Eilks, I., Markic, S. and Ralle, B. (Eds), *Science Education Research and Education for Sustainable Development*, Shaker, Aachen, pp. 157-168.
- Charamba, E. (2022), "Leveraging multilingualism to support science education through translanguaging pedagogy", in Jakobsson, A., Nygård Larsson, P. and Karlsson, A. (Eds), *Translanguaging in Science Education*, Springer, Cham, pp. 257-275.
- Cheung, K.K.C., Pun, J.K.H. and Li, W. (2024), "Students' holistic reading of socio-scientific texts on climate change in a ChatGPT scenario", *Research in Science Education*, doi: [10.1007/s11165-024-10177-2](https://doi.org/10.1007/s11165-024-10177-2).
- Chiu, T.K.F., Ahmad, Z., Ismailov, M. and Sanusi, I.T. (2024), "What are artificial intelligence literacy and competency? A comprehensive framework to support them", *Computers and Education Open*, Vol. 6, p. 100171.
- Chu-Ke, C. and Dong, Y. (2024), "Misinformation and literacies in the era of generative artificial intelligence: a brief overview and a call for future research", *Emerging Media*, Vol. 2 No. 1, pp. 70-85.

-
- Cohen, J. (1960), "A coefficient of agreement for nominal scales", *Educational and Psychological Measurement*, Vol. 20 No. 1, pp. 37-46.
- Cohen, J. (1988), *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed., Lawrence Earlham Associates, Hillsdale.
- Cohen, L., Manion, L. and Morrison, K. (2018), *Research Methods in Education*, 8th ed., Routledge, Abingdon.
- Cooper, G. (2023), "Examining science education in ChatGPT", *Journal of Science Education and Technology*, Vol. 32 No. 3, pp. 444-452.
- Cummins, J. (2017), "Teaching for transfer in multilingual school contexts", in García, O., Lin, A.M.Y. and May, S. (Eds), *Encyclopedia of Language and Education: Vol. 5. Bilingual and Multilingual Education*, 3rd ed., Springer, Cham, pp. 103-115.
- Davies, M. and Barnett, R. (2015), "Introduction", in Davies, M. and Barnett, R. (Eds), *The Palgrave Handbook of Critical Thinking in Higher Education*, Palgrave Macmillan, New York, pp. 1-25.
- DeAndres-Tame, I., Tolosana, R., Vera-Rodriguez, R., Morales, A., Fierrez, J. and Ortega-Garcia, J. (2024), "How good is ChatGPT at face biometrics? A first look into recognition, soft biometrics, and explainability", *IEEE Access*, Vol. 12, pp. 34390-34401.
- Eaton, S.E. (Ed.). (2024), *Second Handbook of Academic Integrity*, Springer, Cham.
- Erduran, S., Guilfoyle, L. and Park, W. (2022), "Science and religious education teachers' views of argumentation and its teaching", *Research in Science Education*, Vol. 52 No. 2, pp. 655-673.
- Flesch, R. (1948), "A new readability yardstick", *Journal of Applied Psychology*, Vol. 32 No. 3, pp. 221-233.
- Furtak, E.M., Heredia, S.C. and Morrison, D. (2019), "Formative assessment in science education", in Andrade, H., Cizek, G. and Bennett, R.E. (Eds), *Handbook of Formative Assessment in the Disciplines*, Routledge, New York, pp. 97-125.
- García, O. (2009), *Bilingual Education in the 21st Century*, Wiley-Blackwell, Malden.
- Gardner-Chloros, P. (2009), *Code-Switching*, Cambridge University Press, Cambridge.
- Gates, B. (2023), "The age of AI has begun", available at: www.gatesnotes.com/The-Age-of-AI-Has-Begun
- George, D. and Mallery, P. (2022), *IBM SPSS Statistics 27 Step by Step*, 17th ed. Routledge, New York.
- Gorrepati, P.L. and Smith, G.P. (2021), "Contrasting readability and availability of Spanish language with English language patient education materials", *Pediatric Dermatology*, Vol. 38 No. S2, pp. 142-143.
- Greening, C. and Lithgow, T. (2020), "Formation and function of bacterial organelles", *Nature Reviews Microbiology*, Vol. 18 No. 12, pp. 677-689.
- Gulati, V., Roy, S.G., Moawad, A., Garcia, D., Babu, A., Poot, J.D. and Teytelboym, O.M. (2024), "Transcending language barriers: can ChatGPT be the key to enhancing multilingual accessibility in healthcare?", *Journal of the American College of Radiology*, doi: [10.1016/j.jacr.2024.05.009](https://doi.org/10.1016/j.jacr.2024.05.009).
- Idsardi, R.C., Luft, J.A., Wingfield, J.L., Whitt, B., Barriga, P.A. and Lang, J.D. (2023), "Relationships between undergraduate instructors' conceptions of how students learn and their instructional practices", *Journal of Research in Science Teaching*, Vol. 60 No. 9, pp. 2076-2110.
- Khalifa, A.A. and Ibrahim, M.A. (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*, doi: [10.1108/AGJSR-09-2023-0423](https://doi.org/10.1108/AGJSR-09-2023-0423).
- Khowaja, S.A., Khuwaja, P., Dev, K., Wang, W. and Nkenyereye, L. (2024), "ChatGPT needs SPADE (sustainability, privacy, digital divide, and ethics) evaluation: a review", *Cognitive Computation*, Vol. 16 No. 5, doi: [10.1007/s12559-024-10285-1](https://doi.org/10.1007/s12559-024-10285-1).
- Kong, S.C., Korte, S.M., Burton, S., Keskitalo, P., Turunen, T., Smith, D., Wang, L., Lee, J.C.-K. and Beaton, M.C. (2024), "Artificial intelligence (AI) literacy—an argument for AI literacy in

- education”, *Innovations in Education and Teaching International*, doi: [10.1080/14703297.2024.2332744](https://doi.org/10.1080/14703297.2024.2332744).
- 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”, *PLOS Digital Health*, Vol. 2 No. 2, p. e0000198.
- Łabuz, M. and Nehring, C. (2024), “Information apocalypse or overblown fears—what AI mis- and disinformation is all about? Shifting away from technology toward human reactions”, *Politics and Policy*, doi: [10.1111/polp.12617](https://doi.org/10.1111/polp.12617).
- Langin, K. (2024), “Researchers buy citations to inflate metrics”, *Science*, Vol. 383 No. 6685, p. 807.
- Lee, H. (2023), “The rise of ChatGPT”, *Anatomical Sciences Education*, Vol. 17 No. 5, pp. 926-931, doi: [10.1002/ase.2270](https://doi.org/10.1002/ase.2270).
- Lemmi, C. and Pérez, G. (2024), “Translanguaging in elementary science”, *International Journal of Science Education*, Vol. 46 No. 1, pp. 1-27.
- Lim, W.M., Gunasekara, A., Pallant, J.L., Pallant, J.I. and Pechenkina, E. (2023), “Generative AI and the future of education”, *The International Journal of Management Education*, Vol. 21 No. 2, pp. 1-13.
- Lo, C.K., Hew, K.F. and Jong, M.S.Y. (2024), “The influence of ChatGPT on student engagement: a systematic review and future research agenda”, *Computers and Education*, Vol. 219, pp. 1-21.
- Long, D. and Magerko, B. (2020), “What is AI literacy?”, Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, *Honolulu*, pp. 1-16.
- Lu, Q., Yao, Y., Xiao, L., Yuan, M., Wang, J. and Zhu, X. (2024), “Can ChatGPT effectively complement teacher assessment of undergraduate students’ academic writing?”, *Assessment and Evaluation in Higher Education*, Vol. 49 No. 5, doi: [10.1080/02602938.2024.2301722](https://doi.org/10.1080/02602938.2024.2301722).
- Marques, N., Silva, R.R. and Bernardino, J. (2024), “Using ChatGPT in software requirements engineering: a comprehensive review”, *Future Internet*, Vol. 16 No. 6, pp. 1-21.
- Martin, P.P. and Graulich, N. (2024), “Beyond language barriers: allowing multiple languages in postsecondary chemistry classes through multilingual machine learning”, *Journal of Science Education and Technology*, Vol. 33 No. 3, pp. 333-348.
- Matsushita, K. (2018), “An invitation to deep active learning”, in Matsushita, K. (Ed.), *Deep Active Learning*, Springer, Singapore, pp. 15-33.
- Mathee, M. and Turpin, M. (2019), “Teaching critical thinking, problem solving, and design thinking”, *Journal of Information Systems Education*, Vol. 30 No. 4, pp. 242-252.
- Mayo, O. (2008), “A century of hardy–weinberg equilibrium”, *Twin Research and Human Genetics*, Vol. 11 No. 3, pp. 249-256.
- Mazak, C.M. and Herbas-Donoso, C. (2015), “Translanguaging practices at a bilingual university”, *International Journal of Bilingual Education and Bilingualism*, Vol. 18 No. 6, pp. 698-714.
- Meyer, A., Ruthard, J. and Streichert, T. (2024), “Dear ChatGPT – can you teach me how to program an app for laboratory medicine?”, *Journal of Laboratory Medicine*, Vol. 0 No. 0, doi: [10.1515/labmed-2024-0034](https://doi.org/10.1515/labmed-2024-0034).
- Mihalache, A., Huang, R.S., Popovic, M.M. and Muni, R.H. (2024), “ChatGPT-4: an assessment of an upgraded artificial intelligence chatbot in the United States medical licensing examination”, *Medical Teacher*, Vol. 46 No. 3, pp. 366-372.
- Millar, A.H., Whelan, J., Soole, K.L. and Day, D.A. (2011), “Organization and regulation of mitochondrial respiration in plants”, *Annual Review of Plant Biology*, Vol. 62 No. 1, pp. 79-104.
- Mizokami, S. (2018), “Deep active learning from the perspective of active learning theory”, in: Matsushita, K. (Ed.), *Deep Active Learning*, Springer, Singapore, pp. 79-91.
- Mogavi, R.H., Deng, C., Kim, J.J., Zhou, P., Kwon, Y.D., Metwally, A.H.S., Tlili, A., Bassanelli, S., Bucchiarone, A., Gujar, S., Nacke, L.E. and Hui, P. (2024), “ChatGPT in education: a blessing or

-
- a curse? A qualitative study exploring early adopters' utilization and perceptions", *Computers in Human Behavior: Artificial Humans*, Vol. 2 No. 1, pp. 1-20.
- Moorhouse, B.L., Yeo, M.A. and Wan, Y. (2023), "Generative AI tools and assessment: guidelines of the world's top-ranking universities", *Computers and Education Open*, Vol. 5, pp. 1-10.
- Naddaf, M. (2023), "ChatGPT generates fake data set to support scientific hypothesis", *Nature*, Vol. 623 No. 7989, pp. 895-896.
- Ng, D.T.K., Tan, C.W. and Leung, J.K.L. (2024), "Empowering student self-regulated learning and science education through ChatGPT: a pioneering pilot study", *British Journal of Educational Technology*, Vol. 55 No. 4, pp. 1328-1353.
- Ng, D.T.K., Lee, M., Tan, R.J.Y., Hu, X., Downie, J.S. and Chu, S.K.W. (2023), "A review of AI teaching and learning from 2000 to 2020", *Education and Information Technologies*, Vol. 28 No. 7, pp. 8445-8501.
- Nieminen, J.H. and Carless, D. (2023), "Feedback literacy: a critical review of an emerging concept", *Higher Education*, Vol. 85 No. 6, pp. 1381-1400.
- Nieuwenhuis, B.P.S., Billiard, S., Vuilleumier, S., Petit, E., Hood, M.E. and Giraud, T. (2013), "Evolution of uni- and bifactorial sexual compatibility systems in fungi", *Heredity*, Vol. 111 No. 6, pp. 445-455.
- OECD (2020), *PISA 2024 Strategic Vision and Direction for Science*, OECD Publishing, Paris.
- OpenAI (2022), "Introducing ChatGPT", available at: <https://openai.com/blog/chatgpt>
- OpenAI (2023), "Introducing ChatGPT Plus", available at: <https://openai.com/blog/chatgpt-plus>
- Osborne, J. and Pimentel, D. (2022), "Science, misinformation, and the role of education", *Science*, Vol. 378 No. 6617, pp. 246-248.
- Osborne, J. and Pimentel, D. (2023), "Science education in an age of misinformation", *Science Education*, Vol. 107 No. 3, pp. 553-571.
- Oxford English Dictionary (OED) (2024), "Online version", available at: www.oed.com, Oxford University Press.
- Piller, I., Zhang, J. and Li, J. (2022), "Peripheral multilingual scholars confronting epistemic exclusion in global academic knowledge production", *Multilingua*, Vol. 41 No. 6, pp. 639-662.
- Polat, H., Topuz, A.C., Yildiz, M., Taslibeyaz, E. and Kursun, E. (2024), "A bibliometric analysis of research on ChatGPT in education", *International Journal of Technology in Education*, Vol. 7 No. 1, pp. 59-85.
- Pretorius, L. (2023), "Fostering AI literacy", *Journal of Academic Language and Learning*, Vol. 17 No. 1, pp. 1-8.
- Pretorius, L. and Cahusac de Caux, B. (2024), *The AI Literacy Framework for Higher Education: A Grounded Theory Exploration of the Foundational, Social, Conceptual, Ethical, and Affective Domains of AI Literacy*, Monash University, Melbourne, doi: [10.26180/25965178.v2](https://doi.org/10.26180/25965178.v2).
- Price, M., Rust, C., O'Donovan, B., Handley, K. and Bryant, R. (2012), *Assessment Literacy: The Foundation for Improving Student Learning*, Oxford Centre for Staff and Learning Development, Oxford.
- Ramnarain, U., Dlamini, T., Bansal, G. and Dhurumraj, T. (2022), "Life sciences teachers' practices of informal formative assessment in inquiry-based teaching", *International Journal of Science Education*, Vol. 44 No. 18, pp. 2745-2762.
- Raymann, K., Brochier-Armanet, C. and Gribaldo, S. (2015), "The two-domain tree of life is linked to a new root for the archaea", *Proceedings of the National Academy of Sciences*, Vol. 112 No. 21, pp. 6670-6675.
- Soo, R.M., Hemp, J., Parks, D.H., Fischer, W.W. and Hugenholtz, P. (2017), "On the origins of oxygenic photosynthesis and aerobic respiration in cyanobacteria", *Science*, Vol. 355 No. 6332, pp. 1436-1440.

- Stajich, J.E., Berbee, M.L., Blackwell, M., Hibbett, D.S., James, T.Y., Spatafora, J.W. and Taylor, J.W. (2009), "The fungi", *Current Biology*, Vol. 19 No. 18, pp. R840-R845.
- Stavans, A., Seroussi, B. and Zadunaisky-Ehrlich, S. (2019), "Literacy-related abilities' effects on argumentative text quality structure", *Journal of Literacy Research*, Vol. 51 No. 3, pp. 315-335.
- Stevens, T.J. and Arkin, I.T. (2000), "Do more complex organisms have a greater proportion of membrane proteins in their genomes?", *Proteins: Structure, Function, and Genetics*, Vol. 39 No. 4, pp. 417-420.
- Stockmeyer, N.O. (2009), "Using Microsoft Word's readability program", *Michigan Bar Journal*, Vol. 88 No. 1, pp. 46-47.
- Tan, C.N.-L., Tee, M. and Koay, K.Y. (2024), "Discovering students' continuous intentions to use ChatGPT in higher education: a tale of two theories", *Asian Education and Development Studies*, doi: [10.1108/AEDS-04-2024-0096](https://doi.org/10.1108/AEDS-04-2024-0096).
- Thorp, H.H. (2023), "ChatGPT is fun, but not an author", *Science*, Vol. 379 No. 6630, p. 313.
- Topping, K.J. (2018), *Using Peer Assessment to Inspire Reflection and Learning*, Routledge, New York.
- UNESCO (2023), *ChatGPT and Artificial Intelligence in Higher Education*, UNESCO.
- Wieman, C. (2017), *Improving How Universities Teach Science*, Harvard University Press, Cambridge.
- Winstone, N. and Carless, D. (2020), *Designing Effective Feedback Processes in Higher Education*, Routledge, Abingdon.
- Wu, Y. and Schunn, C.D. (2023), "Passive, active, and constructive engagement with peer feedback", *Contemporary Educational Psychology*, Vol. 73, pp. 1-14.
- Young, A.K., White, B.T. and Skurtu, T. (2013), "Teaching undergraduate students to draw phylogenetic trees", *Evolution: Education and Outreach*, Vol. 6 No. 1, pp. 1-15.
- Zhu, S., Wang, Z., Zhuang, Y., Jiang, Y., Guo, M., Zhang, X. and Gao, Z. (2024), "Exploring the impact of ChatGPT on art creation and collaboration: benefits, challenges and ethical implications", *Telematics and Informatics Reports*, Vol. 14, pp. 1-12.
- Zonjić, M. (2024), "Need a policy for using ChatGPT in the classroom? Try asking students", *Nature*, doi: [10.1038/d41586-024-01691-4](https://doi.org/10.1038/d41586-024-01691-4).

Further reading

- Chong, S.W. (2021), "Reconsidering student feedback literacy from an ecological perspective", *Assessment and Evaluation in Higher Education*, Vol. 46 No. 1, pp. 92-104.
- Nature (2023), "Tools such as ChatGPT threaten transparent science; here are our ground rules for their use", *Nature*, Vol. 613 No. 7945, p. 612.

Corresponding author

Pablo Antonio Archila can be contacted at: pabloarchila@usta.edu.co

Appendix 1. Argumentative texts generated by ChatGPT**Text 1**

Instructor's prompt: write an argumentative text on the importance of Hardy–Weinberg equilibrium for evolution.

ChatGPT output:

Hardy–Weinberg equilibrium is a fundamental concept in evolutionary theory because it provides a framework for understanding how genetic populations remain stable over time and how they can change over generations.

This equilibrium is based on the idea that the frequency of an allele in a population remains constant from generation to generation, as long as there are no factors that affect the distribution of alleles, such as natural selection, mutation or migration. If a population is in Hardy–Weinberg equilibrium, then the frequency of alleles in the next generation can be predicted with a high degree of accuracy.

In reality, however, populations are rarely in Hardy–Weinberg equilibrium due to the presence of factors that affect the distribution of alleles. Natural selection, for example, can increase the frequency of alleles that confer an adaptive advantage, while mutation can introduce new alleles into the population.

Understanding the Hardy–Weinberg equilibrium is important for understanding how evolution takes place in populations. By understanding how factors such as natural selection and mutation affect the distribution of alleles, one can predict changes in the genetic composition of a population and how these changes may affect its long-term adaptation and survival.

In conclusion, Hardy–Weinberg equilibrium is a key concept for understanding the evolution of populations, as it provides a framework for predicting and explaining changes in their genetic composition. Therefore, it is an essential aspect for the research and study of evolution.

Text 2

Instructor's prompt: write an argumentative text on the importance of Hardy–Weinberg equilibrium for evolution.

ChatGPT output:

Hardy–Weinberg equilibrium is a fundamental concept in genetics and evolution, which describes the distribution of gene frequencies in a population at a given time. This equilibrium is important because it provides a theoretical basis for understanding how genetic and environmental factors affect the allele frequency and genetic structure of a population over time.

The Hardy–Weinberg equilibrium states that, under certain ideal conditions, gene frequencies in a population will remain constant from generation to generation. However, in reality, many factors can interfere with this equilibrium, including natural selection, migration, mutation, and chance. When these factors affect the frequency of alleles in a population, evolution occurs.

Therefore, Hardy–Weinberg equilibrium is a valuable tool for studying evolution, as it allows us to identify perturbations in gene frequency and to determine which factors may be causing changes in the population. In addition, it also allows us to estimate the probability of a particular genetic event occurring in a population, which is useful for predicting the occurrence of genetic diseases and other inherited disorders.

In conclusion, Hardy–Weinberg equilibrium is a key concept in genetics and evolution, providing a sound theoretical basis for understanding how genetic and environmental factors affect the genetic structure of a population. This understanding is essential for predicting and preventing hereditary disorders, and for better understanding how species evolve over time.

Text 3

Instructor's prompt: write an argumentative text comparing the life cycles of plants, animals, protists, and fungi with a dikaryotic phase.

ChatGPT output:

The life cycles of plants, animals, protists, and fungi with a dikaryotic phase share a fundamental characteristic: the presence of two haploid nuclei that fuse to give rise to a diploid nucleus. However, there are significant differences in the way these life cycles develop in each of these groups of organisms.

In plants with a dikaryotic phase, the life cycle is characterized by alternating generations, i.e., alternating between a haploid and a diploid phase. During the haploid phase, the plant produces haploid gametes that fuse to give rise to a diploid cell, which divides and gives rise to a sporophyte, a diploid multicellular structure that produces haploid spores. These spores germinate and give rise to a gametophyte, a haploid multicellular structure that produces gametes. This cycle repeats continuously and is characteristic of flowering plants and non-flowering plants, such as mosses and ferns.

In animals with a dikaryotic phase, the life cycle is characterized by the production of haploid gametes that fuse to give rise to a diploid zygote, which divides and gives rise to a multicellular diploid organism. This organism goes through a process of growth and development until it reaches sexual maturity and produces haploid gametes, thus completing the cycle.

In protists with a dikaryotic phase, the life cycle is also characterized by the production of haploid gametes that fuse to give rise to a diploid cell. This diploid cell may divide by mitosis to produce diploid daughter cells or may divide by meiosis to produce haploid daughter cells that differentiate into gametes. Sexual reproduction in protists is facultative, meaning that it can occur under both favorable and unfavorable conditions.

Finally, in fungi with a dikaryotic phase, the life cycle is characterized by the production of haploid spores that germinate and give rise to a multicellular haploid fungus. This haploid fungus may fuse with another haploid fungus to form a dikaryotic cell that produces a reproductive structure, such as a fruiting body, which in turn produces diploid spores. These diploid spores can germinate and give rise to a haploid fungus, thus completing the cycle.

In summary, although the life cycles of plants, animals, protists and fungi with a dikaryotic phase share a fundamental characteristic, each of these groups of organisms has developed its own life cycle adapted to its environment and reproductive needs. Understanding the differences between these life cycles is essential for understanding biodiversity and the evolution of living things, as well as for protecting and conserving endangered species.

Text 4

Instructor's prompt: write an argumentative text comparing the life cycles of plants, animals, protists, and fungi with a dikaryotic phase.

ChatGPT output:

The life cycles of plants, animals, protists and fungi with a dikaryotic phase are different from those without this phase. The dikaryotic phase is one in which the nuclei of the cell are haploid and fuse to form a diploid nucleus, which will give rise to the diploid phase of the cell.

In plants with a dikaryotic phase, such as liverworts and angiosperms, the diploid phase is found in the reproductive tissue of the plant, such as flowers and fruits. In these plants, pollen is transported from the male haploid phase to the female haploid phase, where fusion of the haploid nuclei and formation of the diploid phase occurs.

In animals with a dikaryotic phase, such as fungi and protists, the dikaryotic phase is necessary for sexual reproduction. In fungi, the dikaryotic phase is found in the fruiting body and is responsible for the production of spores, which will give rise to a new generation of fungi. In protists, the dikaryotic phase occurs during sexual reproduction, which can take place either in the aquatic environment or in the soil.

In the case of fungi, the dikaryotic phase is essential for spore production and sexual reproduction. Fungi reproduce asexually through the production of haploid spores, but sexual reproduction requires the fusion of two haploid nuclei to produce the dikaryotic phase, which in turn will produce the diploid spores.

In summary, the dikaryotic phase is essential for sexual reproduction in organisms that exhibit it, such as plants, animals, protists and fungi. The presence of this phase allows the fusion of haploid nuclei to produce the diploid phase, which will lead to the formation of gametes and spores. Understanding the life cycles of these organisms with a dikaryotic phase is essential to understand biodiversity and the evolution of living beings, and to be able to conserve and protect these endangered species.

Text 5

Instructor's prompt: write an argumentative text about how prokaryotic and eukaryotic organisms obtain energy.

ChatGPT output:

The process by which organisms obtain energy is crucial for their survival, growth, and reproduction. Prokaryotic and eukaryotic organisms differ in their cellular structure and organization, which leads to differences in the ways they obtain energy.

Prokaryotic organisms, such as bacteria, obtain energy through various metabolic pathways, including aerobic and anaerobic respiration, fermentation, and photosynthesis. These pathways involve the breakdown of organic molecules, such as glucose, to release energy in the form of ATP (adenosine triphosphate). In aerobic respiration, the organic molecules are broken down in the presence of oxygen, while in anaerobic respiration, the breakdown occurs in the absence of oxygen. Fermentation is another way in which prokaryotic organisms obtain energy, but it is less efficient than respiration as it only yields a small amount of ATP. In contrast, photosynthetic prokaryotes, such as cyanobacteria, obtain energy by using light energy to produce organic molecules through photosynthesis.

Eukaryotic organisms, such as plants and animals, obtain energy through a process called cellular respiration, which occurs in the mitochondria of their cells. Cellular respiration involves the breakdown of glucose and other organic molecules to produce ATP through a series of metabolic pathways. Unlike prokaryotes, eukaryotic organisms cannot carry out photosynthesis to obtain energy, and thus rely on consuming other organisms or organic matter to obtain nutrients and energy. Plants, on the other hand, are an exception to this rule as they are able to carry out photosynthesis through specialized organelles called chloroplasts.

In conclusion, the ways in which prokaryotic and eukaryotic organisms obtain energy are vastly different. Prokaryotes are able to obtain energy through a variety of metabolic pathways, including photosynthesis, while eukaryotes rely on cellular respiration to break down organic molecules for energy. The different strategies used by these organisms reflect their unique cellular structures and the environments in which they evolved.

Text 6

Instructor's prompt: write an argumentative text about how prokaryotic and eukaryotic organisms obtain energy.

ChatGPT output:

All living organisms require energy to carry out their essential functions, including growth, reproduction, and movement. Prokaryotic and eukaryotic organisms are two types of cells that differ in many ways, including how they obtain energy.

Prokaryotic organisms, such as bacteria and archaea, obtain energy through various processes, including photosynthesis, respiration, and fermentation. Photosynthetic prokaryotes use sunlight to synthesize organic molecules, such as glucose, from carbon dioxide and water. Non-photosynthetic prokaryotes obtain energy by breaking down organic molecules, such as glucose, through respiration or fermentation. Respiration is a more efficient process that uses oxygen to produce more ATP (adenosine triphosphate), the molecule that cells use as their primary energy source. Fermentation, on the other hand, is a less efficient process that does not require oxygen and produces less ATP.

Eukaryotic organisms, including animals, plants, fungi, and protists, obtain energy through a variety of processes, including photosynthesis, respiration, and fermentation. Photosynthetic eukaryotes, such as plants and algae, use sunlight to synthesize organic molecules from carbon dioxide and water. Non-photosynthetic eukaryotes obtain energy by breaking down organic molecules, such as glucose, through respiration or fermentation. Respiration is a more efficient process that uses oxygen to produce more ATP, while fermentation is a less efficient process that produces less ATP.

However, eukaryotic organisms have a unique feature that allows them to obtain energy in a different way than prokaryotic organisms. Eukaryotes have specialized organelles called mitochondria that are responsible for producing ATP through respiration. Mitochondria are believed to have evolved from an ancient prokaryotic organism that was engulfed by a eukaryotic cell. The symbiotic relationship between the two cells allowed the eukaryotic cell to obtain energy more efficiently, leading to the evolution of more complex eukaryotic organisms.

In conclusion, both prokaryotic and eukaryotic organisms obtain energy through various processes, including photosynthesis, respiration, and fermentation. However, eukaryotic organisms have a unique advantage due to the presence of mitochondria, which allows them to produce more ATP through respiration. The evolution of mitochondria in eukaryotic cells is a testament to the adaptability of living organisms and their ability to develop specialized structures to better survive and thrive in their environment.

Text 7

Instructor's prompt: write an argumentative text about the phylogenetic relationships among bacteria, archaea, plants, animals, protists, and fungi.

ChatGPT output:

The phylogenetic relationships among bacteria, archaea, plants, animals, protists, and fungi have long been a topic of debate among scientists. While some have argued that these groups are distinctly separate, recent research has shown that they are more closely related than previously thought.

First, it is important to understand the classification system used to group organisms. Traditionally, organisms have been grouped based on physical characteristics, such as morphology and behavior. However, with the advent of molecular biology, researchers have begun to use genetic information to better understand evolutionary relationships.

Recent studies have used molecular markers such as ribosomal RNA to reconstruct the evolutionary history of life. These studies have shown that bacteria and archaea are more closely related to each other than to eukaryotes, which include plants, animals, protists, and fungi. In fact, some researchers propose that archaea should be classified as a type of bacteria.

Additionally, studies have shown that plants and animals are more closely related to each other than to protists or fungi. This may come as a surprise, as plants and animals appear to be vastly different in terms of their form and function. However, molecular evidence has revealed similarities in their DNA and cellular processes that suggest a shared ancestry.

Protists and fungi, on the other hand, are more distantly related to the other groups. Protists are a diverse group of eukaryotes that do not fit into any other category, while fungi are more closely related to animals than plants. Despite these differences, both protists and fungi play important ecological roles and have significant impacts on human health.

In conclusion, the phylogenetic relationships among bacteria, archaea, plants, animals, protists, and fungi are complex and still not fully understood. However, molecular evidence has revealed that these groups are more closely related than previously thought. This has important implications for our understanding of evolution and the diversity of life on Earth. Ultimately, it highlights the interconnectedness of all living things and the need for continued research to better understand our shared evolutionary history.

Text 8

Instructor's prompt: write an argumentative text about the phylogenetic relationships among bacteria, archaea, plants, animals, protists, and fungi.

ChatGPT output:

The phylogenetic relationships among bacteria, archaea, plants, animals, protists, and fungi have been the subject of intense study and debate for many years. While there is still much to be learned about these relationships, there is a growing body of evidence that supports certain hypotheses.

One of the most widely accepted hypotheses is the Three-Domain System, which divides all living organisms into three major groups: Bacteria, Archaea, and Eukarya. Bacteria and Archaea are both prokaryotes, meaning that they lack a membrane-bound nucleus and other complex cellular structures found in eukaryotes. Eukarya, on the other hand, includes all organisms with a true nucleus and other complex cellular structures.

Within the domain Eukarya, there are four main kingdoms: Plantae, Animalia, Fungi, and Protista. Plantae includes all photosynthetic organisms, such as mosses, ferns, and flowering plants. Animalia includes all animals, from insects and worms to mammals and humans. Fungi includes organisms like mushrooms, yeasts, and molds. Protista is a catch-all kingdom for all eukaryotic organisms that do not fit into the other three kingdoms.

Despite the Three-Domain System being widely accepted, there are still some who argue for different phylogenetic relationships. For example, some researchers argue that the domain Archaea should be split into two separate domains, while others propose that the domain Eukarya should be split into multiple kingdoms.

However, the Three-Domain System has strong support from molecular biology studies, which have shown that bacteria and archaea have significant differences in their genetic makeup, even though they both lack a true nucleus. Additionally, the presence of unique features such as cell walls, membrane lipids, and ribosomal RNA sequences support the separation of the three domains.

In conclusion, while there is still much to be learned about the phylogenetic relationships among bacteria, archaea, plants, animals, protists, and fungi, the Three-Domain System is currently the most widely accepted hypothesis. Molecular biology studies and unique features such as cell walls and membrane lipids provide strong support for this system, and it is likely to continue to be the basis for our understanding of these relationships in the future.

Appendix 2. Survey

1. How often do you use ChatGPT?
 - a. Very frequently
 - b. Fairly frequently
 - c. Infrequently
 - d. Never
2. Apart from this course, have you ever received instruction in the critical examination of texts generated by ChatGPT?
 - a. Yes
 - b. No
3. Do you consider that this course provided you with opportunities to critically evaluate texts generated by ChatGPT?
 - a. Yes
 - b. No

Why?
4. Which part of the written feedback production process was more difficult to develop for you? (more than one option is possible)
 - a. The copy-paste-argue routine related to strengths of the texts generated by ChatGPT
 - b. The copy-paste-argue routine related to weaknesses of the texts generated by ChatGPT
 - c. The copy-paste-edit routine related to weaknesses of the texts generated by ChatGPT

Why?
5. In this course, which type of difficulty did you come up against in assessing the texts generated by ChatGPT in Spanish? (More than one option is possible)
 - a. Anxiety
 - b. Little time to produce written feedback
 - c. Mastery of scientific knowledge
 - d. Scientific vocabulary
 - e. None
 - f. Other
6. In this course, which type of difficulty did you come up against in assessing the texts generated by ChatGPT in English? (More than one option is possible)
 - a. Anxiety
 - b. Grammar
 - c. Little time to produce written feedback
 - d. Mastery of scientific knowledge
 - e. Scientific vocabulary
 - f. None
 - g. Other
7. How often do you have the opportunity to explicitly reflect on the importance of using ChatGPT critically in other university courses?
 - a. Very frequently
 - b. Fairly frequently
 - c. Infrequently
 - d. Never

How far do you agree with the following statements: (1) Strongly disagree, (2) Disagree, (3) Neither agree/ disagree, (4) Agree, and (5) Strongly agree.

8. The activities of critiquing texts generated by ChatGPT helped me become aware of the importance of using this chatbot critically
9. The activities of critiquing texts generated by ChatGPT helped me realize that this chatbot can expose people to false and/or inaccurate scientific information
10. The texts generated by ChatGPT and examined in the activities were related to topics discussed in the Biology of Organisms course
11. The activities of critiquing texts generated by ChatGPT required me to use my prior knowledge
12. The activities of critiquing texts generated by ChatGPT required me to use what I have learned in the Biology of Organisms course
13. The critical revision of texts generated by ChatGPT requires the adoption of healthy skepticism

Source: Authors' own work