

Article

ChatGPT as a Virtual Peer: Enhancing Critical Thinking in Flipped Veterinary Anatomy Education

Nieves Martín-Alguacil *, Luis Avedillo , Rubén A. Mota-Blanco , Mercedes Marañón-Almendros and Miguel Gallego-Agúndez 

Departmental Section of Anatomy and Embryology, School of Veterinary Medicine, Universidad Complutense de Madrid, 28040 Madrid, Spain; luiavedi@ucm.es (L.A.); rubenmot@ucm.es (R.A.M.-B.); mdemaran@ucm.es (M.M.-A.); migugall@ucm.es (M.G.-A.)

* Correspondence: nmartina@ucm.es

Abstract

Artificial intelligence is transforming higher education, particularly in flipped classroom settings, in which students learn independently prior to class and collaborate during in-person sessions. This study examines the role of ChatGPT as a virtual peer in a veterinary anatomy course centered on cardiovascular and respiratory systems. Over two academic years (2023–2025), 297 first-year veterinary students worked in small groups to explore anatomy through structured prompts in English and Spanish using ChatGPT versions 3.5 and 4. Activities involved analyzing AI output, evaluating anatomical accuracy, and suggesting alternative names for vascular variations. Learning outcomes were assessed using Bloom's Taxonomy-based questions, and student perceptions were captured via online surveys. Progressive performance improvement was noted across three instructional phases, particularly in higher-level cognitive tasks (Bloom level 4). Responses to English prompts were more accurate than those to Spanish prompts. While students appreciated ChatGPT's role in reinforcing knowledge and sparking discussion, they also flagged inaccuracies and emphasized the need for critical evaluation. Peer collaboration was found to be more influential than chatbot input. Conclusions: ChatGPT can enrich flipped anatomy instruction when paired with structured guidance. It supports content review, fosters group learning and promotes reflective thinking. However, developing digital literacy and ensuring expert oversight are essential to maximizing the educational value of AI.

Keywords: critical thinking; digital literacy; flipped classroom; generative AI; veterinary anatomy



Academic Editor: Hideki Kasuya

Received: 24 July 2025

Revised: 1 September 2025

Accepted: 1 September 2025

Published: 3 September 2025

Citation: Martín-Alguacil, N.; Avedillo, L.; Mota-Blanco, R.A.; Marañón-Almendros, M.; Gallego-Agúndez, M. ChatGPT as a Virtual Peer: Enhancing Critical Thinking in Flipped Veterinary Anatomy Education. *Int. Med. Educ.* **2025**, *4*, 34. <https://doi.org/10.3390/ime4030034>

Copyright: © 2025 by the authors. Published by MDPI on behalf of the Academic Society for International Medical Education. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The emergence of generative artificial intelligence (GenAI) has triggered a paradigm shift in higher education, particularly in subjects requiring analytical reasoning and conceptual clarity, such as anatomy [1]. ChatGPT, a conversational artificial intelligence (AI), has emerged as a prominent tool in this field due to its ability to simulate academic dialog and facilitate active learning [2]. AI technologies offer students multifaceted support by enabling personalized learning experiences, adaptive feedback and increased accessibility, which particularly benefits those with diverse needs [3–5]. GenAI tools can also improve writing and research skills by helping with idea generation, organization, and language refinement [6–9]. Moreover, by making learning more interactive and responsive, AI can

foster greater motivation and engagement, thereby contributing to better academic outcomes and higher levels of student satisfaction [7,10]. While AI can provide valuable educational support, it also poses risks that may hinder student development. For example, over-reliance on AI tools may result in students accepting outputs without scrutiny or bypassing deeper learning, thereby weakening their critical thinking and problem-solving skills [11,12]. Furthermore, if not carefully evaluated, GenAI misinformation or biased content can lead to academic errors [12]. The ease with which GenAI answers can be accessed also raises concerns about academic integrity, potentially encouraging superficial engagement and reducing students' motivation to learn independently [6,12]. Educators are now faced with the challenge of integrating AI responsibly, striking a balance between technological innovation and academic rigor [13].

In light of the rapid integration of AI into educational settings, it is essential to understand how students adopt and engage with AI tools to promote their effective and ethical use in learning contexts. The integration of AI tools, such as ChatGPT, into educational settings has sparked a growing interest in pedagogical innovation and the adoption of technology. While initial studies have examined the capacity of GenAI to facilitate writing, problem-solving and personalized feedback [14–16], further research is required to comprehend the mechanisms that influence student engagement and learning outcomes when these tools are incorporated into instructional design.

Flipped classroom (FC) models offer fertile ground for such integration. In these settings, students engage with core material independently before class and use class time for applied learning, peer discussion and problem solving [17–20]. When integrated as a virtual peer, ChatGPT can provide students with immediate explanations, conceptual feedback and opportunities for enquiry [21]. However, its presence introduces a novel variable that may not always deliver consistent or factually accurate responses [22]. This raises questions about student trust, digital literacy, and critical engagement in AI-enhanced learning environments [23]. Previous studies have evaluated ChatGPT's ability to respond to biochemistry, physiology and anatomy-related prompts at various cognitive levels [22,24,25]. While the results were promising, discrepancies in response accuracy, especially between languages, highlighted the need for further investigation [22]. For instance, prompts submitted in English tended to produce more reliable, context-aware responses than those in Spanish [22]. Nevertheless, the tool was not immune to hallucinations or superficial reasoning [22], which emphasizes the importance of students' evaluation and verification skills.

This study is framed by the Technology Acceptance Model (TAM) [26,27], which has been widely used in the context of educational technologies. According to TAM, users' adoption of a technology is primarily influenced by their perceptions of its usefulness and ease of use, which shape their attitudes and behavioral intentions. In the context of AI in education, the relevance of TAM has been validated by several studies [28]. Recent research has started to apply TAM to GenAI tools. For example, Dwivedi et al. (2023) [29] suggest that the perceived usefulness of ChatGPT in simplifying academic tasks could boost student motivation and autonomy. Together, these findings suggest that TAM is a robust framework for analyzing student interaction with ChatGPT, particularly in blended or flipped learning environments. Beyond TAM, complementary perspectives are offered by other models such as the Unified Theory of Acceptance and Use of Technology (UTAUT) [30,31] (and the Theory of Planned Behavior (TPB) [32]. UTAUT introduces constructs such as social influence and facilitating conditions, which could be relevant in classroom settings where peer collaboration and instructor guidance influence the use of technology. TPB emphasizes attitudes, subjective norms, and perceived behavioral control, which could inform future studies on ethical considerations and self-regulation in AI-assisted learning. Despite these

theoretical advances, few empirical studies have examined the specific impact of ChatGPT on learning outcomes in structured educational interventions. This study addresses this gap by evaluating the impact of ChatGPT-supported flipped learning on student performance and retention, while also exploring how TAM constructs manifest in real-world classroom dynamics. Building on earlier research, this study moves beyond static chatbot testing to an immersive classroom application. Conducted within a veterinary anatomy course focusing on cardiovascular and respiratory systems, the study introduces ChatGPT as an active participant in group learning sessions. The aim was to observe how students interacted with the AI, how they interpreted its contributions, and how they balanced its input against that of their peers and experts. Through structured exercises, including evaluating anatomical patterns, vascular variations, and reviewing AI-generated responses, students navigated increasingly complex cognitive tasks aligned with Bloom's taxonomy (Bloom, 1956) [33].

Therefore, this study aims to understand the pedagogical potential of ChatGPT in flipped learning environments, as well as the cognitive and ethical challenges it presents. By observing how students navigate the chatbot's presence in collaborative learning, this research will contribute to the wider discussion about the role of AI in higher education and the skills students require to interact with it responsibly and effectively. To assess the impact on learning, student performance was evaluated through the regular course exam using questions designed across four cognitive levels based on Bloom's taxonomy [34], enabling a structured analysis of both knowledge acquisition and the development of key academic skills. The integration of GenAI into health science education is changing the way students approach complex subjects such as anatomy. As new tools such as ChatGPT enter the classroom, educators must balance innovation with academic rigor.

2. Materials and Methods

This study formed part of the core first-year subject of the Veterinary Medicine degree program, Anatomy and Embryology I, and focused specifically on the respiratory and cardiovascular systems. Data was collected across two academic years (2023/24 and 2024/25), the study was embedded within a FC framework. Both groups were taught by the same instructors using identical materials and assessments.

The study content was presented using H5P (HTML5 Package), an open-source tool that enables the creation and sharing of rich, interactive HTML5 content, including videos, presentations, quizzes, and games.

The Wooclap platform was used to collect data on students' experiences. Wooclap is an interactive platform designed to boost student engagement during live or remote sessions. It enables educators to incorporate real-time polls, quizzes, word clouds and other interactive features into their presentations, thereby encouraging active participation and facilitating immediate feedback.

Face to face class groups organization. Students were divided into discussion groups of approximately six students each. In 2023/24 ChatGPT version used was GPT-3.5. and in the 2024/25 academic year ChatGPT version GPT-4. In both academic years, ChatGPT was introduced as a virtual participant in student discussion groups during in-class sessions. Accessible via a digital interface, the chatbot interacted with students in real time, responding to prompts related to the cardiovascular and respiratory systems. Students could engage with the chatbot in either Spanish or English. They were also encouraged to formulate questions, or discussion prompts for ChatGPT in either language, depending on their preference. The study monitored how the language of the prompt influenced the quality, relevance and accuracy of the chatbot's responses. To explore its educational value, ChatGPT was embedded into student discussion groups, guiding learners through a

three-phase progression—from critical thinking to anatomical reasoning, and ultimately, the application of that reasoning in real-world scenarios.

All participants had a similar academic background and equal access to course materials, including lecture notes, textbooks and institutional resources. Prior to the intervention, students completed a baseline survey to collect information on their academic history and study habits. This information was used to confirm group assignment and AI engagement status. To promote peer-level interaction, some discussion groups were formed exclusively of repeating students. This grouping strategy was implemented to encourage balanced participation from people with similar academic backgrounds and experience of the course content.

For the outcomes assessment, participants were divided into two groups:

1. **Experimental Group:** Students who attended face-to-face FC sessions in which AI tools (specifically ChatGPT versions 3.5 and 4) were integrated into the learning process. These students were instructed to use ChatGPT exclusively during in-class activities and not outside designated sessions.
2. **Control Group:** Students who did not attend any face-to-face FC sessions. Based on self-reported data from the initial survey, these students did not engage with ChatGPT or any other AI tools during their studies. They prepared for the course solely using traditional resources such as textbooks, lecture notes and peer discussions.

To ensure clear separation between the groups, AI usage was strictly limited to the classroom environment for the experimental group. As they did not attend the sessions, the control group had no exposure to AI tools and confirmed in the initial survey that they did not use AI for studying. Both groups received access to the same core curriculum and learning objectives. The experimental group engaged with AI-assisted activities designed to enhance their understanding of anatomical and embryological concepts through guided prompts and cognitive-level questioning. The control group studied independently using conventional methods.

Instructional Phases and Cognitive Assessment:

- The FC intervention was structured into three progressive instructional phases, each of which targeted specific cognitive processes aligned with Bloom's taxonomy.
 - Phase I: Focused on foundational knowledge and critical evaluation, with the vomeronasal organ (VNO) serving as the topic. This was assessed through cognitive level 1 (recall) and cognitive level 3 (application) questions.
 - Phase II focused on anatomical reasoning involving persistent ductus arteriosus (PDA) and persistent truncus arteriosus (PTA). This was assessed through cognitive level 1 and cognitive level 3 questions, which evaluated both basic understanding and applied reasoning.
 - Phase III addressed problem solving and comparative anatomy using the vascular pattern of the aortic arch. This was assessed through cognitive level 2 (comprehension) and cognitive level 4 (analysis) questions, which measure deeper conceptual integration.

Interactive H5P videos were used to support pre-class learning and scaffold understanding of these processes. Each phase was designed to progressively increase the cognitive engagement of students: Phase I focused on discrimination and critical thinking, requiring students to identify relevant anatomical data and evaluate the accuracy of ChatGPT-generated responses. Phase II emphasized anatomical reasoning, prompting students to critique flawed AI explanations and refine their understanding of congenital cardiovascular anomalies. Phase III involved problem solving, where students applied their comparative anatomical knowledge to solve a complex problem.

Outcome measures:

Student performance was evaluated using a set of cognitive-level questions tailored to each phase. Additionally, a general comparison was made using four standard cognitive-level questions and one GPT-specific cognitive-level-3 question, as presented in Table 1. (a) Cognitive level 1, knowledge (e.g., recalling anatomical structures); (b) Cognitive level 2, comprehension (e.g., diagram or image interpretation); (c) Cognitive level 3, application (e.g., association of concepts, anatomical reasoning); and (d) Cognitive level 4, analysis (e.g., anatomical reasoning to solve real-life problems, anatomy and diagnostic imaging). These questions were integrated into formal assessments related to the cardiovascular and respiratory systems and were used to measure students' understanding of anatomy and their ability to apply it in context. These assessments enabled comparative analysis of the experimental and control groups' performance in relation to general and phase-specific learning outcomes.

Table 1. Student performance in the regular exam, broken down by cognitive level and including the average score for the ChatGPT-related question.

Academic Year			MEAN SCORE	CL 1	CL 2	CL 3	CL 3 GPT	CL 4
2023/24	<i>n</i> = 151	EL <i>n</i> = 26	4.36 ± 2.62	3.61 ± 3.64	4.42 ± 4.47	3.84 ± 3.13	4.55 ± 2.53	5.55 ± 3.68
		FL <i>n</i> = 125	5.21 ± 2.91	4.39 ± 3.84	5.06 ± 4.08	4.97 ± 3.65	5.19 ± 2.17	6.40 ± 3.65
			* <i>p</i> = 0.024	<i>p</i> = 0.24	<i>p</i> = 0.4	* <i>p</i> = 0.04	<i>p</i> = 0.52	<i>p</i> = 0.15
2024/25	<i>n</i> = 139	EL <i>n</i> = 38	4.89 ± 2.05	4.30 ± 3.52	5.91 ± 3.12	4.45 ± 2.14	3.79 ± 2.99	5.13 ± 3.17
		FL <i>n</i> = 101	5.48 ± 2.44	5.38 ± 4.45	6.36 ± 2.98	4.21 ± 2.05	5.11 ± 3.07	5.95 ± 3.17
			<i>p</i> = 0.26	<i>p</i> = 0.25	<i>p</i> = 0.53	<i>p</i> = 0.65	* <i>p</i> = 0.015	<i>p</i> = 0.25

n = 139, 36 students were excluded from the table due to completing the course through continuous assessment without sitting the final exam. CL: Cognitive level; EL: Expositive learning; FL: Flipped learning; GPT: ChatGPT question (cognitive level 3); *n*: number of students. This table presents the mean scores (± SD) of students from two academic years (2023/24 and 2024/25), comparing expositive learning (control group) and flipped learning (FL group) across four cognitive levels (CL1 to CL4), including a GPT-specific CL3 question. The number of students per group is indicated, and statistical significance was assessed using the Mann–Whitney U test due to non-normal data distribution. Asterisks (*) denote *p*-values < 0.05. In 2024/25, 36 students were excluded from analysis due to completing the course via continuous assessment without sitting the final exam. CL: Cognitive Level; AA: Active Learning; GPT: ChatGPT-supported question.

The cognitive-level questions were integrated into the regular course assessments and remained consistent across cohorts. GPT-generated questions were clearly identified within the exam structure. Although the assessments were not reviewed by independent educators, one exam was examined by a reclamation committee following a student appeal. This committee lacked formal pedagogical training, and their review emphasized the importance of involving qualified educators in evaluating active learning outcomes.

Two main types of data were collected: (1) Student feedback: This was collected through Wooclap interactive sessions and Virtual Campus (VC) surveys. This focused on students' perceptions of ChatGPT's usefulness, credibility and role in group discussions. (2) Learning performance: Academic performance data were collected to assess potential correlations between ChatGPT integration and student outcomes. Performance metrics were drawn from assessments related to the cardiovascular and respiratory systems block. The sample sizes for the performance analysis were *n* = 151 (2023/24) and *n* = 139 (2024/25). During the 2024/25 academic year, 36 students were exempted from assessment after fulfilling the requirements of the course through continuous evaluation, without taking the final examination.

Conceptual framework:

To gain a better understanding of the factors that influence students' engagement with ChatGPT in a flipped learning environment, this study adopts the TAM as its guiding theoretical framework. Originally developed by Davis (1989) [26], the TAM has been widely

used to explain the adoption of new technologies by users in various domains, including education. The model posits that two primary beliefs—perceived usefulness (PU) and perceived ease of use (PEOU)—shape users' attitudes towards a technology, which in turn influence their behavioral intention to use it and ultimately their actual use. In the context of this study, TAM provides a lens through which to examine how students perceive and interact with ChatGPT as a learning support tool. Specifically:

- Perceived usefulness refers to students' belief that ChatGPT enhances their academic performance, supports comprehension and facilitates task completion.
- Perceived ease of use captures students' perception of ChatGPT as an intuitive, accessible and user-friendly tool.

These perceptions inform students' attitude towards use, which influences their intention to use ChatGPT in future learning scenarios. Actual use is reflected in the frequency and depth of ChatGPT engagement during the course.

By incorporating the TAM into the research design, this study evaluates the effectiveness of ChatGPT in enhancing learning outcomes and explores the psychological and behavioral mechanisms that underlie its adoption. This framework allows for a more nuanced interpretation of the results and contextualizes the findings within the wider discourse on the adoption of educational technology. Furthermore, the use of TAM aligns with prior research on AI integration in education, providing a validated structure for analyzing student behavior and informing future pedagogical strategies. Thus, the conceptual framework serves as both a theoretical foundation and a practical guide for interpreting the implications of ChatGPT use in higher education.

The study employed both qualitative and quantitative methods. Qualitative analysis was used to identify themes related to trust, engagement and the perceived value of ChatGPT in student feedback, while quantitative analysis was used to assess academic performance.

Procedure:

A. Initial and final surveys.

The initial surveys were conducted during the briefing session, before active learning began, and the final surveys were administered at the end of the experience. The Woodclap platform was used to collect data on students' experiences of using AI and ChatGPT.

A thematic analysis of open-ended survey responses using a deductive approach based on the TAM was conducted. Initial codes were developed based on the TAM constructs of perceived usefulness, perceived ease of use, attitude towards use and behavioral intention, and were refined through iterative reading of the data. Two researchers coded the responses independently and resolved any discrepancies through discussion to ensure consistency. Themes were then synthesized, and representative quotes were selected to illustrate each category.

B. Study content.

The study was structured into three phases, each of which was preceded by an interactive H5P video to support learning before the class. A distinct cognitive exercise was selected for each of the three phases of the study to target specific learning objectives and assess the corresponding cognitive processes.

Phase I: The VNO.

Phase II: PDA and PTA.

Phase III: The vascular pattern of the aortic arch.

C. Cognitive exercise.

During the in-person sessions, students took part in a two-part cognitive exercise designed to reinforce and apply the content introduced in the flipped video. First, each student completed the activity individually on the Wooclap platform. They were then asked to repeat the exercise in small groups, recording their answers on paper and discussing them with their peers to compare reasoning and conclusions. Phase I began with a Wooclap quiz focused on identifying the anatomical location of the VNO. Phase II involved a Wooclap activity in which students identified a patent ductus arteriosus in an anatomical image and answered the following question: How does this condition affect intracardiac, pulmonary and systemic circulation? Phase III began with a written test in which students had to identify the vascular pattern of the aortic arch across various domestic species, including the major associated vessels. The transition from individual to group work in all phases was intended to promote deeper understanding through collaborative reasoning and discussion.

D. Chatbot interaction.

ChatGPT was consulted in three progressive phases, each with a specific learning objective. Phase I: The goal was to develop students' critical thinking and information discrimination skills. After completing the Wooclap exercise, each group selected an animal species (horse, dog or rabbit) and asked ChatGPT to provide information on the location of the VNO in that species. Phase II: The objective was to foster anatomical reasoning. Students were prompted to ask ChatGPT about the difference between PDA and PTA, reasoning through the anatomical distinctions. Phase III: The objective was to apply anatomical reasoning to a clinical scenario. After taking a written test on the vascular pattern of the aortic arch in domestic species, the students asked ChatGPT, 'What is the bovine aortic arch?'

E. Use of real anatomical images.

Real anatomical images were used to support the exercises. Phase I: Anatomical cross-sections of the horse, dog and rabbit. Phase II: prosections of hearts with PDA and PTA. Phase III: dissections of the aortic arch in various domestic species. These visual materials enabled students to validate their anatomical reasoning and reinforce their understanding through direct observation.

F. Data analysis.

Data were analyzed using non-parametric statistical methods, given that the distributions of the test scores did not meet criteria for normality according to the Kolmogorov–Smirnov test. For each academic year and cognitive level, mean scores and standard deviations were computed for both the reference group (students not attending class and following the traditional methodology) and the study group (students attending class and using active learning). Comparative analyses between groups were performed using the Mann–Whitney U test. Statistical significance was considered at $p < 0.05$. All analyses were conducted using [Stata: Release 15_Statistical Software]. Results are presented as mean \pm standard deviation, and p -values are reported for each comparison.

3. Results

A. Initial and final surveys.

To explore students' evolving engagement with AI technologies, we administered a survey via Wooclap over two academic years: 2023/24 and 2024/25. We interpreted the findings through the lens of the TAM, which suggests that perceived usefulness and ease of use are key drivers of technology adoption.

Adoption trends and familiarity. AI tool usage increased from 51% (71 out of 138) to 86% (116 out of 135), suggesting a rapid normalization of AI in academic contexts. This shift may reflect increased exposure, as well as institutional or peer-driven encouragement. ChatGPT dominance grew from 27% (35 out of 128) to 62% (87 out of 140), suggesting not only broader AI adoption, but also a shift towards a single dominant tool. This may be due to ChatGPT's user-friendly interface and perceived usefulness, which aligns with TAM's "perceived ease of use" dimension. Familiarity gains: The drop in the proportion of students who were unfamiliar with ChatGPT, falling from 5% (8 out of 156) to 3% (4 out of 135), and those who had heard of it but had never used it, falling from 54% (85 out of 156) to 10% (13 out of 135) suggests an improvement in digital literacy. This supports the idea that students are moving from passive awareness to active engagement.

Usage patterns and functional integration. Information retrieval saw a threefold increase from 22% (34 out of 156) to 64% (86 out of 135), indicating a shift in how students conduct academic research. This suggests that ChatGPT is increasingly being viewed as a viable alternative to traditional search engines. Translation and text editing usage also rose significantly, indicating a diversification of use cases. These functions fall under the 'augmentation' level of the SAMR model, where technology enhances existing tasks. Assignment support usage increased from 13% (20 out of 156) to 62% (84 out of 135), raising pedagogical and ethical questions. While this reflects its perceived usefulness, it also highlights the need for clearer guidelines on its responsible use. Dishonest use doubled, albeit from a low base 3% (5 out of 156) to 7% (9 out of 135). This modest increase signals the need for academic integrity policies to evolve alongside technological capabilities.

Perceptions of educational value. Positive perceptions of ChatGPT as a learning tool increased from 49% (61 out of 124) to 71% (95 out of 133), indicating a growing alignment between student requirements and the capabilities of AI. This finding is consistent with TAM's "perceived usefulness" construct. Skepticism declined slightly from 16% (20 out of 124) to 11% (15 out of 133), indicating that concerns about AI's educational value are diminishing, possibly due to increased exposure and peer validation. Critical literacy remains high, with 90% of students recognizing the importance of evaluating AI-generated content. This is encouraging, as it suggests that students are not blindly accepting outputs, but are developing metacognitive strategies to assess credibility.

B. Cognitive Exercise.

Some GPT responses contained inaccuracies. Although 85% of students correctly identified the location of the VNO in the initial Woodclap activity, all groups initially accepted the incorrect response provided by ChatGPT. The chatbot's answer was inaccurate in terms of both anatomical terminology and content. A similar pattern emerged in subsequent phases. For example, when analyzing heart prosections displaying a PDA and a PTA, and when identifying the aortic arch pattern in various domestic species, students once again accepted the chatbot's incorrect information, despite having previously studied the material.

Using real anatomical materials, such as cross-sectional images of the nasal cavities in horses, dogs and rabbits (Phase I) (Figure 1), heart prosections showing PDA and PTA (Phase II) (Figure 2) and dissections of the aortic arch in different species (Phase III), allowed students to verify anatomical structures and reasoning. These resources supported the validation of their group responses and provided a basis for comparison with the chatbot's output.

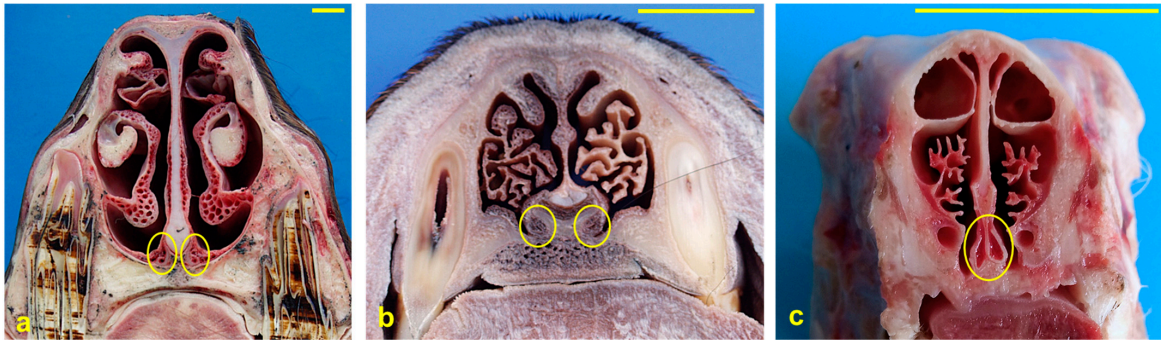


Figure 1. This real anatomical model was used to solve the Phase I ChatGPT cognitive exercise, which shows cross-sections of the middle third of the nasal cavities of a horse (a), a dog (b) and a rabbit (c). The location of the vomeronasal organ (VNO) (circled yellow in each case) and its relationship with the nasal cavity can be seen clearly. Scale bar: 3 cm.

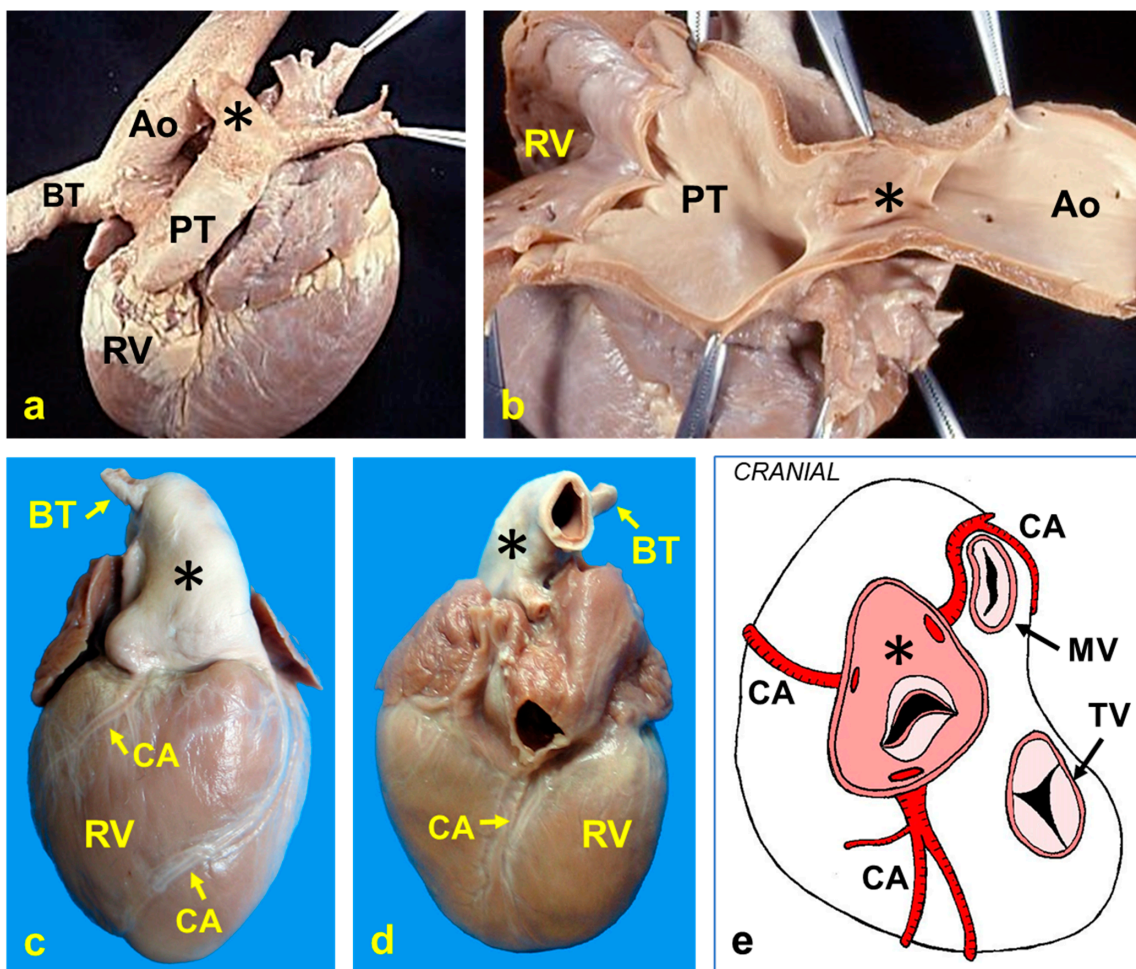


Figure 2. Heart dissections showing PDA and PTA were used to solve the Phase II ChatGPT cognitive exercise. (a) Left dorsolateral view of the heart of a calf with a patent ductus arteriosus (asterisk), which communicates the pulmonary trunk with the aorta (Ao), distal to the point of origin of the brachiocephalic trunk. (b) Anatomical demonstration of the patent communication (asterisk) between the aorta and the pulmonary trunk in the same heart. (c) Cranial view and (d) caudal view of the heart of a newborn piglet affected by PTA (asterisk), which first and anomalously emits three coronary arteries before distally emitting the brachiocephalic trunk. (e) Semi-schematic drawing of the cardiac base of the same piglet's heart to facilitate understanding of the distribution of the truncus arteriosus (asterisk). Ao: aorta; BT: brachiocephalic trunk; CA: coronary artery; MV: mitral valve; PT: pulmonary trunk; RV: right ventricle; TV: tricuspid valve.

Performance was assessed in the regular exam session using four different cognitive level questions. A comparative analysis of the academic years 2023/24 and 2024/25 is presented in Table 1. When the GPT-generated questions from the two academic years analyzed are compared, a slight improvement in performance is observed (Table 1).

The effectiveness of ChatGPT as a learning tool was evaluated in three progressive phases throughout the 2024/25 academic year. Questions aligned with the different cognitive levels of Bloom's Taxonomy were used for this evaluation. In Phases I and II, Levels 1 (e.g., recalling anatomical structures) and 3 (e.g., concept association and anatomical reasoning) were applied. Phase III used Levels 2: Comprehension (e.g., interpreting diagrams or images) and 4: Analysis (e.g., applying anatomical reasoning to real-life problems or integrating anatomy with diagnostic imaging). The results of this assessment are presented in Table 2.

Table 2. Student performance on ChatGPT-related questions in the regular exam, organized by the three instructional phases.

PHASE		CL 1	CL 2	CL 3	CL 4
I	EL <i>n</i> = 38	4.45 ± 3.17		3.13 ± 3.53	
	FL <i>n</i> = 101	5.78 ± 3.36 * <i>p</i> = 0.029		4.44 ± 3.49 * <i>p</i> = 0.034	
II	EL <i>n</i> = 38	4.02 ± 3.59		2.42 ± 2.36	
	FL <i>n</i> = 101	5.54 ± 4.48 <i>p</i> = 0.099		2.49 ± 2.24 <i>p</i> = 0.88	
III	EL <i>n</i> = 38		5.64 ± 3.05		5.29 ± 3.48
	FL <i>n</i> = 101		6.45 ± 2.97 <i>p</i> = 0.2		5.91 ± 3.53 <i>p</i> = 0.36

CL: Cognitive level; EL: Expositive learning; FL: Flipped learning; GPT: ChatGPT question (cognitive level 3); *n*: number of students. This table shows the mean scores (± SD) of students across three instructional phases (I, II and III), comparing traditional and active learning groups. Each phase corresponds to a different stage of GPT integration into peer discussion groups within a flipped learning model. Performance is evaluated at various cognitive levels (CL1 to CL4) and statistical significance is assessed using the Mann–Whitney U test. Asterisks (*) indicate *p*-values <0.05. CL: Cognitive Level; AA: Active Learning; GPT: ChatGPT-supported instruction.

In Phase I, students used ChatGPT primarily as an information source, with the goal of practicing discrimination and critical thinking—identifying relevant data, evaluating the accuracy of responses, and distinguishing between reliable facts and potential AI-generated inaccuracies. To conclude the activity, the students examined real anatomical specimens to identify the location of the VNO in the three selected species (Figure 1). This step laid the foundation for deeper anatomical reasoning in the following phases.

In Phase II, students used ChatGPT to explain complex anatomical concepts. When asked to differentiate between PDA and PTA, the chatbot's response misused terms and failed to clarify the distinction—giving students a chance to practice anatomical reasoning and spot errors in AI-generated content. After interacting with ChatGPT, students individually reflected on its responses, identifying strengths and weaknesses. Then, through group discussion, each team crafted an anatomical justification for their answer and critically analyzed the chatbot's output—building skills in both reasoning and collaborative evaluation. To finish the activity, students looked at real anatomical specimens showing two heart conditions: PDA and PTA (Figure 2).

In Phase III, students used anatomical reasoning to tackle clinical cases—specifically applying their knowledge of the comparative aortic arch vascular pattern. This helped them connect anatomical concepts with real-world diagnostic thinking and clinical imaging. The initial response from ChatGPT did not include any connection between the bovine aortic arch and the vascular variation in human anatomy. This gap highlighted the need to

refine the prompt to guide the chatbot toward more clinically relevant details—showing how careful prompt design is key when applying AI to medical reasoning. To deepen their clinical understanding, students were encouraged to use digital tools like Google Search or AI sources to explore the anatomy of the bovine aortic arch. This step promoted independent research, critical comparison of sources, and reflection on how anatomical variations relate to clinical practice. To bring clinical relevance into focus, students were shown an actual angiographic image where a bovine aortic arch was identified in a human patient. They were encouraged to explore it using digital tools—including Google Lens—to investigate anatomical patterns, interpret imaging features, and connect veterinary concepts to human cardiovascular variation. In the final group task, students analyzed the bovine aortic arch pattern and compared it with different animal species aortic arch vascular pattern. Their goal was to find which animal has a similar vascular layout and explain the match based on anatomy. Once they made their choice, they were invited to name the variation creatively—replacing “bovine” with a species they felt was anatomically closest. For example, if they picked a dog, they could rename it the “canine arch.” This exercise encouraged anatomical reasoning, creativity, and clinical thinking—all in one go.

In terms of performance assessment across the three phases, questions in phases I and II were designed at cognitive levels 1 and 3, respectively. No difference in memorization (cognitive level 1) was noted between phases I and II, but an improvement in the formulation of anatomical reasoning (cognitive level 3) was evident. In Phase III, the questions were designed at Cognitive Levels 2 and 4, and a significant improvement in performance was evident (Table 2). When the average scores across all three phases are compared, a progressive improvement is evident. When comparing student performance on cognitive level 3 questions—those requiring reasoning and concept association—results showed a noticeable improvement in the questions worked with ChatGPT. Across the two academic years, students scored better on GPT-supported tasks than on similar-level questions in the standard exam, suggesting that AI integration can enhance deep learning when paired with guided instruction.

4. Discussion

AI tools such as ChatGPT show great potential in education, supporting the development of dynamic and engaging learning environments [7,10,13,35,36]. Although the FC model has been widely adopted across disciplines, it presents specific challenges in veterinary anatomy, given the subject’s focus on spatial reasoning, memorizing complex structures and integrating theory with practical dissection and imaging [17–19,37]. This study examined the potential of ChatGPT to support learning at four cognitive levels, ranging from basic recall to diagnostic reasoning [22,34]. It also investigated the integration of ChatGPT into a FC model for anatomy education, a subject renowned for its high cognitive demand and content density. While ChatGPT provided useful content, it occasionally produced incorrect information, highlighting the importance of human oversight [22,38–40]. GPT-4 was found to be more accurate and relevant than GPT-3.5, though both still required careful review. Despite its limitations, ChatGPT can effectively enhance engagement and personalized learning in FC settings when used judiciously [22].

Unlike discussion-based subjects, anatomy requires a scaffolded, multimodal approach involving 3D visuals, interactive quizzes, and video demonstrations in order to manage cognitive load and prepare students for practical sessions. Therefore, adapting the flipped model to these tactile and visual demands is essential, with studies showing that it can significantly enhance comprehension and engagement compared to traditional formats. Using anatomical cross-sections, such as VNO images of various species and heart or aortic arch dissections, helped students to validate their reasoning with visual evidence and to

correct ChatGPT's errors through observation and discussion. Tasks such as comparing vascular patterns, identifying variations and analyzing angiographic images encouraged creativity and clinical relevance.

The 2024/25 study followed a three-phase instructional model aligned with Bloom's Taxonomy. Each phase was preceded by an H5P video to activate prior knowledge and promote independent learning. Phase I emphasized digital literacy and critical thinking in relation to the VNO, Phase II focused on anatomical reasoning regarding congenital cardiovascular anomalies, and Phase III involved problem solving regarding bovine aortic arch vascular variation. Students engaged with ChatGPT throughout, evaluating its responses and identifying errors, thereby reinforcing critical analysis and contextual understanding. Despite occasional inaccuracies, the structured use of AI promoted deeper learning and maintained scientific rigor through peer collaboration and systematic design. While many students correctly answered anatomical questions, they often accepted ChatGPT's incorrect responses without scrutiny, a trend observed throughout all phases. This highlights the importance of strengthening digital literacy and critical thinking, particularly in AI-enhanced learning environments [41]. Student performance improved across the three instructional phases, in line with the increasingly complex cognitive tasks. While Phases I and II, which focused on memorization and reasoning (levels 1 and 3), produced modest gains, this suggests that foundational exposure to tools such as ChatGPT supports recall, but additional scaffolding is required for deeper thinking. In Phase II (Cognitive Level 3), the performance outcome was notably lower than in the other phases, suggesting that the instructional approach used may have been less effective in supporting students' development of intermediate reasoning skills. This finding warrants further investigation, particularly about the timing and structure of GPT integration during this phase. Phase III, which targeted levels 2 and 4, demonstrated stronger improvement through comprehension and problem solving. Students often identified inaccuracies in ChatGPT's responses, which reinforced the importance of critical appraisal and well-structured prompts [42]. Instructor-led group discussions enhanced students' ability to articulate and refine their reasoning, particularly in Phase III, where they applied anatomical knowledge to clinical scenarios. When combined with instructor guidance and collaborative reasoning, AI has the potential to enhance flipped anatomy education by fostering a deeper understanding and encouraging creativity [43].

In this study, GPT was not used as a source of content delivery, but rather as a didactic tool to stimulate peer discussion and promote critical thinking. This aligns with findings that GenAI can improve writing and reasoning skills by facilitating idea generation and language refinement [6–9]. Our results from Phase I suggest that GPT-supported sessions may enhance performance at lower cognitive levels, where a solid understanding of the fundamentals is paramount. Furthermore, the interactive nature of GPT appeared to foster student engagement, which is consistent with reports of increased motivation and satisfaction in AI-enhanced learning environments [7,10]. However, our study also reflects the complexities and limitations associated with AI use in education. As noted in the literature, over-reliance on GenAI tools can lead students to accept outputs uncritically, potentially bypassing deeper cognitive processing [11,12]. This is a particular concern at higher cognitive levels (CL3 and CL4), where our results showed no significant differences between groups, likely due to shared exposure to active learning in practical sessions overshadowing the specific impact of GPT. While statistical significance is a valuable metric for identifying measurable differences between groups, it should not be the sole criterion for interpreting educational outcomes—particularly in the context of cognitive-level assessments. Questions aligned with Bloom's taxonomy aim to capture complex learning processes such as reasoning, problem-solving, and critical thinking, which may

not always produce large numerical differences or statistically significant results. In our study, the goal was not to demonstrate superiority through statistical thresholds, but to explore how students engaged with AI-supported learning environments and how tools like ChatGPT influenced their cognitive development. Even when differences in scores were not statistically significant, observable trends—such as improved performance in lower-level tasks or increased engagement—offered meaningful pedagogical insights. Moreover, the multifactorial nature of educational settings, including prior knowledge, motivation, and exposure to active learning strategies, makes it difficult to isolate the impact of a single intervention. Therefore, we interpret statistical outcomes alongside qualitative observations and instructional context to provide a more holistic understanding of learning progression. Additionally, the potential for misinformation or biased content generated by GenAI tools remains a critical issue [12]. In our instructional design, students were encouraged to challenge and verify GPT-generated responses in order to mitigate these risks and reinforce analytical skills. Nevertheless, the ease with which AI-generated answers can be accessed raises valid concerns about academic integrity and independent learning [6,12], and these concerns must be carefully considered in future implementations.

Over the two academic years, student feedback indicated growing comfort with ChatGPT, highlighting its potential as a learning aid [44]. Most students expressed a positive attitude towards AI-assisted learning, with an increased academic use of AI for tasks such as information retrieval, text editing and assignment support in 2024/25 [45,46]. A small minority admitted to dishonest use, underscoring the need for ethical guidelines. Reflections revealed mixed views: while many students valued ChatGPT for discussion and clarifying concepts, others noted confusion or overreliance [47]. Concerns about the accuracy of GenAI highlight the importance of distinguishing between AI-generated content and verified knowledge, and of fostering digital literacy in medical education [47,48]. Peer collaboration remained a preferred learning method, suggesting that AI should enhance, rather than replace, human interaction. In line with the TAM, students who found ChatGPT useful and easy to use were more likely to adopt it; its interface supports engagement and collaborative learning [49].

While the use of AI tools in medical education is gaining momentum, our findings emphasize the potential and challenges of such integration. One of the primary challenges was the difficulty in isolating ChatGPT's impact due to students' preference for peer interaction. This confounding variable suggests that, although AI can support learning, human collaboration remains a dominant and valued component of educational experience. Additionally, the study did not systematically validate the accuracy of AI-generated content or assess students' baseline digital literacy—two factors that could significantly influence learning outcomes. These omissions limit the interpretability of the results, emphasizing the need for future research to incorporate robust validation protocols and digital competency assessments. One of the main limitations of this study is the absence of a fully isolated control group that did not engage with ChatGPT at any stage. Although we attempted to mitigate this issue by including comparison groups, such as students who did not attend face-to-face flipped sessions or GPT-supported discussions, the fact that all students were exposed to mandatory practical sessions based on Team-Based Learning (TBL) complicates the interpretation of the results. As these sessions are designed to foster reasoning and problem-solving skills and are attended by all students regardless of group, they likely contributed to similar performance at higher cognitive levels (CL4). Consequently, it is challenging to attribute differences in learning outcomes exclusively to the AI intervention. Furthermore, this study was not designed to evaluate knowledge acquisition from ChatGPT itself, but rather to explore its role in promoting discussion, critical thinking and information discrimination as a didactic tool. Future research should employ more

controlled experimental designs to better isolate the impact of AI tools and investigate their long-term effects on cognitive development.

Another limitation was the lack of pedagogical expertise in the assessment review process. Assessing higher-order skills requires moving beyond traditional correction methods and adopting pedagogical training that supports analysis, synthesis and evaluation. As an example, one exam was reviewed by a reclamation committee, the absence of instructional training among its members emphasized the importance of involving educators with a background in pedagogy in the evaluation of learning outcomes. This is particularly important when assessing active learning, as a nuanced understanding of instructional design and student engagement is essential. Future iterations of this study will therefore include independent educational reviewers to enhance the reliability and validity of assessments.

There is a wealth of literature documenting concerns about over-reliance on GenAI tools such as ChatGPT [11,12,50]. Critics argue that excessive dependence on AI may lead to diminished critical thinking and passive learning. However, this study proactively addressed these concerns by designing learning activities that required students to evaluate, contrast and reflect on the accuracy of AI-generated anatomical content. Rather than passively accepting AI responses, students were encouraged to question and critique the information provided, thereby fostering deeper engagement and independent reasoning. This approach is consistent with constructivist learning theories, which emphasize active knowledge construction through inquiry and reflection.

Within the FC framework, ChatGPT was not positioned as a replacement for human insight, but rather as a catalyst for inquiry and peer discussion. By integrating AI into structured collaborative learning activities, the study mitigated the risk of cognitive passivity, transforming a potential limitation into an opportunity to foster critical thinking. This pedagogical strategy fosters metacognitive abilities and empowers learners to take charge of their learning journey. AI tools such as ChatGPT may cause students to accept output uncritically, particularly when responses are phrased with high confidence despite containing inaccuracies or 'hallucinations'. Our findings echo previous studies' concerns about the persuasive nature of AI-generated misinformation, emphasizing the importance of structured evaluation tasks in counteracting this tendency.

Language discrepancies were identified, noting that English prompts consistently yielded more accurate and coherent responses than Spanish ones. This underlines the importance of prompt quality and linguistic precision in shaping AI output. To mitigate these issues, we emphasize the role of educators in fostering digital literacy and critical appraisal skills. We suggest incorporating scaffolded activities that require students to compare AI-generated content with verified sources, reflect on inconsistencies and engage in peer-led discussions. Furthermore, we recommend professional development for instructors to enable them to design AI-integrated tasks that encourage enquiry, skepticism, and ethical awareness. These strategies aim to position AI not as a shortcut, but as a tool for cultivating deeper learning and the responsible use of technology [3,4].

The results of this study advocate for the wider adoption of AI-supported flipped learning models, particularly in subjects such as anatomy, where the complexity of the content can impede student comprehension. When integrated into well-designed instructional environments, tools like ChatGPT can enhance student autonomy, engagement, and academic performance [3,4]. These findings are consistent with recent studies demonstrating improved learning outcomes and student satisfaction in veterinary anatomy through blended learning strategies [37,51–53]. Furthermore, AI integration enables individualized pacing and a deeper conceptual understanding. This allows students to engage with the material at their own level and revisit complex topics as required [3,4,7,10]. This flexibility

is a vital aspect of contemporary pedagogical frameworks that prioritize learner-centered approaches and adaptive instruction.

To ensure the meaningful and responsible implementation of AI in education, institutions must invest in faculty development, digital infrastructure, and ethical guidelines [54,55]. Educators require training in both the technical use of AI tools and in designing pedagogically sound activities that utilize AI to augment, rather than replace human learning. Ethical considerations, including data privacy, algorithmic bias, and equitable access, must also be addressed to ensure that the integration of AI benefits all learners.

In conclusion, while challenges remain, this study demonstrates that AI tools such as ChatGPT can be effectively integrated into FC models to promote active learning, critical thinking, and student engagement. By positioning AI as a collaborative partner rather than a passive content provider, educators can harness its potential to enrich the learning experience. Future research should continue to explore cross-disciplinary applications, the long-term effects on learning, and strategies for scaling up AI-enhanced pedagogy in diverse educational contexts.

Author Contributions: Conceptualization, N.M.-A.; methodology, N.M.-A., L.A., R.A.M.-B., M.M.-A. and M.G.-A.; validation, N.M.-A., L.A. and R.A.M.-B.; formal analysis, N.M.-A., L.A. and R.A.M.-B.; investigation, N.M.-A., L.A. and R.A.M.-B.; resources, N.M.-A., L.A., R.A.M.-B., M.M.-A. and M.G.-A.; data curation, N.M.-A., L.A., R.A.M.-B., M.M.-A. and M.G.-A.; writing—original draft preparation, N.M.-A.; writing—review and editing, N.M.-A., L.A. and R.A.M.-B.; supervision, N.M.-A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Following consultation with the Bioethics Committee of the Universidad Complutense de Madrid (UCM), it was concluded that this study did not require formal ethical approval. The research was conducted as part of an educational innovation initiative and involved no clinical procedures or collection of sensitive personal data.

Informed Consent Statement: All participating students were informed in advance about the purpose and nature of the study, and their participation was entirely voluntary. Data were collected anonymously to ensure privacy and confidentiality. The study was conducted in accordance with the ethical principles of transparency, respect and academic integrity.

Data Availability Statement: Data and materials are available on request.

Acknowledgments: The authors would like to thank the first-year veterinary students at Complutense University for their active participation throughout the experience and those who did not wish to participate for providing valuable information for the study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Hersh, W. Generative Artificial Intelligence: Implications for Biomedical and Health Professions Education. *Annu. Rev. Biomed. Data Sci.* **2025**, *8*, 355–380. [[CrossRef](#)] [[PubMed](#)]
2. Mittal, U.; Sai, S.; Chamola, V.; Sangwan, D. A Comprehensive Review on Generative AI for Education. *IEEE Access* **2024**, *12*, 142733–142759. [[CrossRef](#)]
3. Chen, L.; Chen, P.; Lin, Z. Artificial Intelligence in Education: A Review. *IEEE Access* **2020**, *8*, 75264–75278. [[CrossRef](#)]
4. Sajja, R.; Sermet, Y.; Cikmaz, M.; Cwiertny, D.; Demir, I. Artificial Intelligence-Enabled Intelligent Assistant for Personalized and Adaptive Learning in Higher Education. *Information* **2023**, *15*, 596. [[CrossRef](#)]
5. Gordon, M.; Daniel, M.; Ajiboye, A.; Uraiby, H.; Xu, N.Y.; Bartlett, R.; Hanson, J.; Haas, M.; Spadafore, M.; Grafton-Clarke, C.; et al. A Scoping Review of Artificial Intelligence in Medical Education: BEME Guide No. 84. *Med. Teach.* **2024**, *46*, 446–470. [[CrossRef](#)] [[PubMed](#)]

6. Alharbi, W. AI in the Foreign Language Classroom: A Pedagogical Overview of Automated Writing Assistance Tools. *Educ. Res. Int.* **2023**, *2023*, 4253331. [[CrossRef](#)]
7. Song, C.; Song, Y. Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Front. Psychol.* **2023**, *14*, 1260843. [[CrossRef](#)] [[PubMed](#)]
8. Kim, J.; Yu, S.; Detrick, R.; Li, N. Exploring students' perspectives on Generative AI-assisted academic writing. *Educ. Inf. Technol.* **2024**, *30*, 1265–1300. [[CrossRef](#)]
9. Nguyen, A.; Hong, Y.; Dang, B.; Huang, X. Human-AI collaboration patterns in AI-assisted academic writing. *Stud. High. Educ.* **2024**, *49*, 847–864. [[CrossRef](#)]
10. Boubker, O. From chatting to self-educating: Can AI tools boost student learning outcomes? *Expert Syst. Appl.* **2023**, *238*, 121820. [[CrossRef](#)]
11. Darvishi, A.; Khosravi, H.; Sadiq, S.; Gašević, D.; Siemens, G. Impact of AI assistance on student agency. *Comput. Educ.* **2023**, *210*, 104967. [[CrossRef](#)]
12. Zhai, C.; Wibowo, S.; Li, L. The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learn. Environ.* **2024**, *11*, 28. [[CrossRef](#)]
13. Boscardin, C.K.; Gin, B.; Golde, P.B.; Hauer, K.E. ChatGPT and Generative Artificial Intelligence for Medical Education: Potential Impact and Opportunity. *Acad. Med.* **2024**, *99*, 22–27. [[CrossRef](#)]
14. Zawacki-Richter, O.; Marín, V.I.; Bond, M.; Gouverneur, F. Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *Int. J. Educ. Technol. High. Educ.* **2019**, *16*, 39. [[CrossRef](#)]
15. Kasneci, E.; Sessler, K.; Küchemann, S.; Bannert, M.; Dementieva, D.; Fischer, F.; Gasser, U.; Groh, G.; Günemann, S.; Hüllermeier, E.; et al. ChatGPT for good? On opportunities and challenges of large language models for education. *Learn. Individ. Differ.* **2023**, *103*, 102274. [[CrossRef](#)]
16. Moritz, S.; Romeike, B.; Stosch, C.; Tolks, D. Generative AI (gAI) in medical education: Chat-GPT and co. *GMS J. Med. Educ.* **2023**, *40*, Doc54. [[CrossRef](#)] [[PubMed](#)]
17. Martín Alguacil, N.; Avedillo, L.J. Theoretical teaching of veterinary anatomy using the flipped classroom method: Evaluation of student performance and perception. In *La Universidad Innova en Metodologías y Herramientas*; Peter Lang—International Academic Publishers: Berlin, Germany, 2024; pp. 197–208. ISBN 978-3-631-91602-5.
18. Martín-Alguacil, N.; Mota Blanco, R.; Avedillo, L.J. Integración de TIC en una experiencia de aprendizaje invertido (Flipped-Classroom) en la enseñanza teórica. In *III Jornada «Aprendizaje Eficaz con TIC en la UCM»*; Ediciones Complutense; Universidad Complutense de Madrid: Madrid, Spain, 2024; pp. 115–127. ISBN 978-84-669-3856-3.
19. Martín-Alguacil, N.; Avedillo, L.; Gallego-Agúndez, M.; Mota-Blanco, R. Student-centered learning: Some issues and recommendations for its implementation in a traditional curriculum setting in health sciences. *Educ. Sci.* **2024**, *14*, 1179. [[CrossRef](#)]
20. Matthew, S.M.; Schoenfeld-Tacher, R.M.; Danielson, J.A.; Warman, S.M. Flipped classroom use in veterinary education: A multinational survey of faculty experiences. *J. Vet. Med. Educ.* **2019**, *46*, 97–107. [[CrossRef](#)] [[PubMed](#)]
21. Dos Santos, O.L.; Cury, D. Challenging the Confirmation Bias: Using ChatGPT as a Virtual Peer for Peer Instruction in Computer Programming Education. In Proceedings of the 2023 IEEE Frontiers in Education Conference (FIE), College Station, TX, USA, 18–21 October 2023; pp. 1–7. [[CrossRef](#)]
22. Martín-Alguacil, N.; Avedillo, L.J.; Mota Blanco, R.; Gallego-Agúndez, M. Utilización del Chat GPT como Herramienta Didáctica en la Enseñanza Invertida en Ciencias de la Salud. In *III Jornada «Aprendizaje Eficaz con TIC en la UCM»*; Ediciones Complutense; Universidad Complutense de Madrid: Madrid, Spain, 2024; pp. 143–157. ISBN 978-84-669-3856-3.
23. Ahn, S. The Impending Impacts of Large Language Models on Medical Education. *Korean J. Med. Educ.* **2023**, *35*, 103–107. [[CrossRef](#)]
24. Banerjee, A.; Ahmad, A.; Bhalla, P.; Goyal, C. Assessing the Efficacy of ChatGPT in Solving Questions Based on the Core Concepts in Physiology. *Cureus* **2023**, *15*, e43314. [[CrossRef](#)]
25. Ghosh, A.; Bir, A. Evaluating ChatGPT's Ability to Solve Higher-Order Questions on the Competency-Based Medical Education Curriculum in Medical Biochemistry. *Cureus* **2023**, *15*, e37023. [[CrossRef](#)] [[PubMed](#)]
26. Davis, F.D. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Q.* **1989**, *13*, 319–340. [[CrossRef](#)]
27. Davis, F.D.; Bagozzi, R.P.; Warshaw, P.R. User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Manag. Sci.* **1989**, *35*, 982–1003. [[CrossRef](#)]
28. Adli, M.R.S.C.D. A Review of the Acceptance on Learning Management System (LMS) Using Unified Theory Acceptance and Use of Technology (UTAUT). *Int. J. Bus. Entrepren. SMEs* **2023**, *5*, 22–32. [[CrossRef](#)]

29. Dwivedi, Y.K.; Kshetri, N.; Hughes, L.; Slade, E.L.; Jeyaraj, A.; Kar, A.K.; Baabdullah, A.M.; Koohang, A.; Raghavan, V.; Ahuja, M.; et al. "So What If ChatGPT Wrote It?" Multidisciplinary Perspectives on Opportunities, Challenges and Implications of Generative Conversational AI for Research, Practice and Policy. *Int. J. Inf. Manag.* **2023**, *71*, 102642. [[CrossRef](#)]
30. Venkatesh, V.; Morris, M.G.; Davis, G.B.; Davis, F.D. User Acceptance of Information Technology: Toward a Unified View. *MIS Q.* **2003**, *27*, 425–478. [[CrossRef](#)]
31. Venkatesh, V.; Thong, J.Y.L.; Xu, X. Unified Theory of Acceptance and Use of Technology: A Synthesis and the Road Ahead. *J. Assoc. Inf. Syst.* **2016**, *17*, 328–376. [[CrossRef](#)]
32. Ajzen, I. The Theory of Planned Behavior. *Organ. Behav. Hum. Decis. Process.* **1991**, *50*, 179–211. [[CrossRef](#)]
33. Bloom, B.S. (Ed.) *Taxonomy of Educational Objectives, Handbook I: Cognitive Domain*; David McKay Co.: New York, NY, USA, 1956; p. 201.
34. Martín Alguacil, N.; Avedillo, L.J. Student-Centered Active Learning Improves Performance in Solving Higher-Level Cognitive Questions in Health Sciences Education. *Int. Med. Educ.* **2024**, *3*, 346–362. [[CrossRef](#)]
35. Miller, M. ChatGPT, Chatbots and Artificial Intelligence in Education. *Ditch That Textbook 2023*. Available online: <https://ditchthattextbook.com/ai/> (accessed on 22 July 2025).
36. Khan, R.A.; Jawaid, M.; Khan, A.R.; Sajjad, M. ChatGPT—Reshaping Medical Education and Clinical Management. *Pak. J. Med. Sci.* **2023**, *39*, 605–607. [[CrossRef](#)] [[PubMed](#)] [[PubMed Central](#)]
37. de Brito, C.; Terrado, J. Flipping Veterinary Biochemistry, Anatomy, and Physiology: Students' Engagement and Perception. *Vet. Sci.* **2024**, *11*, 354. [[CrossRef](#)] [[PubMed](#)]
38. Bhattacharyya, M.; Miller, V.M.; Bhattacharyya, D.; Miller, L.E. High Rates of Fabricated and Inaccurate References in ChatGPT-Generated Medical Content. *Cureus* **2023**, *15*, e39238. [[CrossRef](#)] [[PubMed](#)]
39. Wagner, M.W.; Ertl-Wagner, B.B. Accuracy of Information and References Using ChatGPT-3 for Retrieval of Clinical Radiological Information. *Can. Assoc. Radiol. J.* **2024**, *75*, 69–73. [[CrossRef](#)] [[PubMed](#)]
40. Lee, S.W.; Choi, W.J. Utilizing ChatGPT in Clinical Research Related to Anesthesiology: A Comprehensive Review of Opportunities and Limitations. *Anesth. Pain Med.* **2023**, *18*, 244–251. [[CrossRef](#)] [[PubMed](#)] [[PubMed Central](#)]
41. Lin, Z. Why and How to Embrace AI Such as ChatGPT in Your Academic Life. *R. Soc. Open Sci.* **2023**, *10*, 230658. [[CrossRef](#)] [[PubMed](#)] [[PubMed Central](#)]
42. Monzon, N.; Hays, F.A. Leveraging Generative Artificial Intelligence to Improve Motivation and Retrieval in Higher Education Learners. *JMIR Med. Educ.* **2025**, *11*, e59210. [[CrossRef](#)] [[PubMed](#)] [[PubMed Central](#)]
43. Singh, K.; Bharatha, A.; Sa, B.; Adams, O.P.; Majumder, M.A.A. Teaching Anatomy Using an Active and Engaging Learning Strategy. *BMC Med. Educ.* **2019**, *19*, 149. [[CrossRef](#)] [[PubMed](#)] [[PubMed Central](#)]
44. Safranek, C.W.; Sidamon-Eristoff, A.E.; Gilson, A.; Chartash, D. The Role of Large Language Models in Medical Education: Applications and Implications. *JMIR Med. Educ.* **2023**, *9*, e50945. [[CrossRef](#)] [[PubMed](#)] [[PubMed Central](#)]
45. Buabbas, A.J.; Miskin, B.; Alnaqi, A.A.; Ayed, A.K.; Shehab, A.A.; Syed-Abdul, S.; Uddin, M. Investigating Students' Perceptions towards Artificial Intelligence in Medical Education. *Healthcare* **2023**, *11*, 1298. [[CrossRef](#)]
46. Straková, N.; Válek, J. Chatbots as a Learning Tool: Artificial Intelligence in Education. *Pedagog. Dipl. II* **2024**, *11* (Suppl. S1), 245–265. [[CrossRef](#)]
47. Reinke, N.B.; Parkinson, A.L.; Kafer, G.R. A Tutorial Activity for Students to Experience Generative Artificial Intelligence: Students' Perceptions and Actions. *Adv. Physiol. Educ.* **2025**, *49*, 461–470. [[CrossRef](#)] [[PubMed](#)]
48. Preiksaitis, C.; Rose, C. Opportunities, Challenges, and Future Directions of Generative Artificial Intelligence in Medical Education: Scoping Review. *JMIR Med. Educ.* **2023**, *9*, e48785. [[CrossRef](#)] [[PubMed](#)] [[PubMed Central](#)]
49. Meyer, J.G.; Urbanowicz, R.J.; Martin, P.C.N.; O'Connor, K.; Li, R.; Peng, P.C.; Bright, T.J.; Tatonetti, N.; Won, K.J.; Gonzalez-Hernandez, G.; et al. ChatGPT and Large Language Models in Academia: Opportunities and Challenges. *BioData Min.* **2023**, *16*, 20. [[CrossRef](#)] [[PubMed](#)]
50. Vieriu, A.M.; Petrea, G. The Impact of Artificial Intelligence (AI) on Students' Academic Development. *Educ. Sci.* **2025**, *15*, 343. [[CrossRef](#)]
51. Gómez, O.; García-Manzanares, M.; Chicharro, D.; Juárez, M.; Llamazares-Martín, C.; Soriano, E.; Terrado, J. Application of Blended Learning to Veterinary Gross Anatomy Practical Sessions: Students' Perceptions of Their Learning Experience and Academic Outcomes. *Animals* **2023**, *13*, 1666. [[CrossRef](#)]
52. Previdelli, R.L.; Boardman, E.; Frill, M.; Frean, S.; Channon, S.B. Supporting Collaborative Dissection through the Development of an Online Wiki Positively Impacts the Learning of Veterinary Anatomy. *Anat. Sci. Educ.* **2024**, *17*, 88–101. [[CrossRef](#)] [[PubMed](#)]
53. Singh, A.A.; Shapter, F.M.; Bernard, A.; Whitworth, D.J.; Holt, M.G.; Waller, P.S.; Bond, S.L. Applying Iterative Student Feedback across Flipped Classroom and Flexible Teaching Approaches: Impact on Veterinary Students' Learning Experience. *Animals* **2024**, *14*, 2335. [[CrossRef](#)] [[PubMed](#)] [[PubMed Central](#)]

54. Kim, J.; Lee, H.; Cho, Y. Learning Design to Support Student-AI Collaboration: Perspectives of Leading Teachers for AI in Education. *Educ. Inf. Technol.* **2022**, *27*, 6069–6104. [[CrossRef](#)]
55. Lo, C.; Hew, K. A Review of Integrating AI-Based Chatbots into Flipped Learning: New Possibilities and Challenges. *Front. Educ.* **2023**, *8*, 1175715. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.