

Exploring ChatGPT's Efficacy in Pathophysiological Analysis: A Comparative Study of Ischemic Heart Disease and Anaphylactic Shock Cases

Alexandra Ina BULBOACĂ^{1,2}, Bogdan BORLEA³, Adriana Elena BULBOACĂ^{4,*}, Ioana STANESCU⁵, and Sorana D. BOLBOACĂ¹

¹ Department of Medical Informatics and Biostatistics, "Iuliu Hațieganu" University of Medicine and Pharmacy, 40349 Cluj-Napoca, Romania

² Clinical Rehabilitation Hospital, 400347 Cluj-Napoca, Romania

³ Faculty of Medicine, "Iuliu Hațieganu" University of Medicine and Pharmacy, 400012 Cluj-Napoca, Romania

⁴ Department of Pathophysiology, "Iuliu Hațieganu" University of Medicine and Pharmacy, 400012 Cluj-Napoca, Romania

⁵ Department of Neurology, "Iuliu Hațieganu" University of Medicine and Pharmacy, 400012 Cluj-Napoca, Romania

⁶ Romanian Society of Medical Informatics, 300222 Timișoara, Romania

E-mails: bulboaca.alexandra@yahoo.com (A.I.B); bogdanborlea27@gmail.com (B.B.);

adriana_bulboaca@yahoo.com (A.E.B.); Ioanastane@yahoo.com (I.S.); sbolboaca@umfcluj.ro (S.D.B.)

* Author to whom correspondence should be addressed;

Received: 2 March 2024/Accepted: 28 March 2024 / Published online: 31 March 2024

Abstract

Physiopathology is one of the basic disciplines which helps to understand the mechanisms of diseases and, accessing effective and personalized therapies. The purpose of this work was to compare the answers provided by human intelligence with experience and professional qualification with the answers provided by ChatGPT, regarding the targeted questions addressed to two clinical cases, to decipher the key pathophysiological mechanisms. Two cases published in an applied pathophysiology book represented the input data. ChatGPT-3.5 was used in our study to retrieve responses to five specific questions per case and GPTZero to score the responses. The input and output text were in Romanian language, and we collected the answers on August 7, 2023. Two researchers independently did the classification of the answers provided by ChatGPT as correct, partially correct, or incorrect. The ChatGPT generated answers were partially correct in 4 out of 5 questions and incorrect in one question in both cases. The number of words generated by ChatGPT was similar to the response given by professionals and the perplexity, burstiness, readability and simplicity have smaller scores. GPTZero did not identify the responses generated by ChatGPT as generated by a large language model. ChatGPT-3.5 generated satisfactory responses but not complete in most of the cases, but must be supervised by experts to ensure the validity of the responses and explanations.

Keywords: ChatGPT; Large Language Model (LLM); Medical cases; Physiopathology

Introduction

Considering the accelerated evolution of the tools for accessing scientific medical information, it is necessary to adapt the medical teaching methods in accordance with these developments. Chat Generative Pre-Trained Transformer (ChatGPT) is a large language model (LLM) instrument with access to a broad spectrum of data including articles, books, and websites. Initiatives are being made towards the aggregation and management of

health data from various sources, including medical facilities, social services, caregivers, and public health. Examples of these initiatives include "Data-based Health Management Initiatives" in Japan, "My Health Record" in Australia, and "NHS Digital" in the United Kingdom. These initiatives serve as valuable sources of healthcare and research data [1-4]. ChatGPT is developed and released by OpenAI and uses unsupervised pre-training and supervised fine-tuning to produce responses similar to those produced by humans. Its algorithm has over 170 billion parameters trained with text available on internet [5,6]. ChatGPT supports multiple languages, including Romania [7], can identify the language and provide answers in that specific language. ChatGPT can be used as a working tool that promptly provides detailed answers to targeted questions, revolutionizing the way medical professionals and general population interact with technology [8,9]. However, it is unlikely that would substitute the human staff in medical health care [10]. ChatGPT training is based on conversation tasks and human questions and can constitute a complex background for knowledge reinforcement. The question formulation and user's interaction lead the ChatGPT responses [11]. Limitations such as inaccurate or inconsistent answers, biases (e.g., in data training reflected in the output, linguistic, ideological, confirmation, temporal, commercial, cognitive, attention, format, source, novelty, availability biases, etc.), privacy and copyright (including inability to understand ethics) had been acknowledged [12-16]. The misuse or abuse of ChatGPT is challenging and the users should be trained for a competent use [5,17]. The role of ChatGPT in higher education is debated, with acknowledgement of quick knowledge assimilation (potential superficial learning) and a by-passing in depth reading and critical analysis and thinking [18,19], which is essential in medicine.

Physiopathology is one of the basic disciplines which helps students to understand the mechanisms of diseases towards effective and personalized therapies. ChatGPT could contribute to a better understanding of pathophysiological mechanisms by improving the medical educational process. The human introduces a collection of questions referring to a specific medical subject, ChatGPT generates the answers, followed by analyzes and correction of the responses by human [20]. Integrating ChatGPT answers in day-by-day activity offers medical information not only for healthcare professionals but also for patients [21]. If healthcare professionals could validate the information provided by ChatGPT, the patients can receive correct and updated medical information for a specific disease. Furthermore, ChatGPT could contribute to doctor-patient relationship and a better patient understanding of the diagnosis and treatment [22]. In clinical surgery, ChatGPT can be used as a useful instrument for writing surgery protocols at the end of the intervention. Although ChatGPT can be an adjuvant tool for medical examination, generating useful and precise questions according with patient's symptoms, the specialist contribution remains essential [23].

Even though ChatGPT offers very rapid responses, there are situations when the generated answer is credible but incorrect. The reliability and accuracy of answers are crucial for medical practice, therefore the ChatGPT answers can confuse those who are asking questions and a validation of the accuracy of the answer is needed [24]. Comprehensive guidelines are still needed to evaluate the accuracy and safety of responses received from ChatGPT towards their implementation in healthcare flow, such as daily practice, patient education, and medical research [25]. For research and educational process, ChatGPT could be a tool for fast access to medical literature and pharmacology databases, offering an up-to-date information [26,27]. ChatGPT can also generate precise and logical questions about a specific content, being useful for medical training and knowledge assessments [28].

The purpose of this work was to compare the answers provided by ChatGPT on one-hand with the answers provided by human intelligence with experience and professional qualification, regarding the targeted questions addressed to some clinical cases, to explore the key pathophysiological mechanisms. In learning pathophysiology, this task is useful in the medical education process because the students must understand the etiology, risk factors, pathogenesis, and diseases complications, based on human normal morpho-functional knowledge.

Materials and Methods

Clinical Cases

Two cases (Case 1 and Case 2) were chosen from the book entitled *Applied Pathophysiology* [29] and were used in this study. For each case, we provided the text, questions and responses given by human professionals, experts

with over 30 years' experience. The human professional's responses to the questions were the gold standard (GS) in the evaluation of the ChatGPT responses.

Case 1. A 61-year-old patient presented to the emergency department (ED) for an acute, intense, morning pain in the precordial region, accompanied by nausea and vomiting. The pain radiates in the upper left limb and the last two fingers, as well as retrosternal. The patient is a smoker and consumes alcohol occasionally, has hypertension controlled with medication and type 2 DM (Diabetes Mellitus), treated with oral antidiabetics, BMI (Body Mass Index) is 31. At the clinical examination, the semiology elements of the pain indicate a cardiac origin. The heart rate was 55 b/min (the patient is on chronic treatment with beta-blockers), the respiratory rate was 22/min, T (body temperature) = 36.8°C, BP (blood pressure) = 90/60 mmHg, and the state of consciousness preserved. The EKG (electrocardiogram) examination revealed typical changes of AMI (Acute Myocardial Infarction), with an elevation of the ST segment in V1-V4, inversion of the T wave in D I, aVL, V2-V5. Laboratory tests: basal blood glucose 185 mg/dL, total CST (cholesterol) 220 mg/dL, LDL-CST 160 mg/dL, TG (triglycerides) 150 mg/dL, HDL-CST 30 mg/dL, leukocytes 10,000/mm³, platelets 450,000/mm³, erythrocytes sedimentation rate (ESR) 20. Cardiac ultrasound reveals decreased LVEF (Left Ventricular Ejection Fraction), without signs of cardiac remodeling. The clinical evolution of the patient was favorable after the treatment.

Questions:

- 1.1. What are the conditions that can cause pain similar to STEMI (ST-elevation AMI) pain?
- 1.2. What are the subclasses of AMI?
- 1.3. Explain the etiopathogenetic mechanisms of AMI in this case.
- 1.4. Indicate the value of cardiac enzymes in the diagnosis of AMI.
- 1.5. List the risk factors/comorbidities that this patient has.

Human professional's answers:

- GS.1.1. Conditions that can simulate pain from painful ischemic heart disease: aortic dissection, cardiomyopathies, aortic stenosis, cardiac ischemia associated with cocaine abuse, etc. There are conditions that are associated with ST elevation on the EKG, but are not accompanied by pain of cardiac origin: left ventricular hypertrophy, left bundle branch block, hyperkalemia, Brugada syndrome. Conditions that can be accompanied by precordial pain but are not associated with cardiac ischemia are represented by pleurisy, pleuritis, costochondritis, pericarditis, gastro-esophageal reflux disease, hiatal hernia, etc. Anamnesis and para-clinical examinations contribute to the differential diagnosis of these conditions.
- GS.1.2. UDMI (Universal Definition of Myocardial Infarction) [30] classifies MI into **type 1 - AMI** that appear because of thrombosis of an atherosclerotic plaque, **type 2 - AMI** caused by to decreased oxygenation of the myocardium from other causes than atherothrombosis, **type 3 - AMI** associated with sudden death, **type 4 - AMI** associated with percutaneous coronary interventions, and **type 5 - AMI** associated with coronary bypass.
- GS.1.3. Etiopathogenetic mechanisms - AMI is associated with the death of myocardial cells, as a consequence of prolonged ischemia. After 10-15 minutes of ischemia, damage to the sarcolemma and mitochondria occurs as a result of the decrease in cellular glucose reserves. Myocytes death progresses from the subendocardium to the subepicardium within hours. The prolongation of myocytes survival depends on the existence and functionality of the collateral circulation and on previous episodes of ischemia/reperfusion that contribute to the phenomenon of "ischemic preconditioning".
- GS.1.4. AMI biomarkers are represented by troponin I (cTnI) and T (cTnT), enzymes found in cardiomyocytes. Increases in cTnI are also associated with progressive muscular dystrophies. Increases in cTn are associated with acute myocardial damage. There are various causes that can induce acute myocardial damage: AMI (atherothrombotic lesions that produce coronary obstruction, myocardial damage by reducing myocardial perfusion from causes other than ATS (coronary spasm, coronary artery emboli, coronary artery dissections, severe prolonged bradycardias, insufficiency breathing, severe anemias), increased heart oxygen needs (prolonged tachyarrhythmias, severe hypertension), myocardial damage associated with other heart conditions (heart failure, cardiomyopathies, interventional cardiology

procedures, electrical defibrillation), or systemic conditions (stroke, pulmonary embolism, pulmonary HT, sepsis, chemotherapy, etc.).

GS.1.5. Risk factors/associated comorbidities - dyslipidemia, smoking, obesity, DM, HT (hypertension) - favor endothelial dysfunctions and atherothrombosis.

Case 2. A 37-year-old patient presented to the ED with a face, neck, thorax and limbs cutaneous rash, palpebral, labial and lingual oedema, persistent emesis, wheezing and pharyngeal foreign body sensation. The patient took an oral Ibuprofen 400 mg pill approximately 25 minutes before presenting to the ED. The heart rate was 120 b/m, the blood pressure was BP=70/45 mmHg and the O₂ blood saturation (SpO₂) 94%.

Questions:

- 2.1. In an anaphylactic reaction, which cells are involved?
- 2.2. What are the mediators of an anaphylactic reaction and what are their pathophysiological results?
- 2.3. Can you explain biphasic anaphylaxis?
- 2.4. What are the vital risk-associated clinical manifestations of an anaphylactic reaction and the pathophysiological principles behind the therapy?
- 2.5. Anaphylactic shock: What are the pathophysiological principles of therapy?

Human professional's answers:

- GS.2.1. The cells involved in causing anaphylaxis are mastocytes and basophils, which are activated through a type 1 hypersensitivity reaction. The allergen binds simultaneously to two specific IgE molecules, anchored to the membrane of the mastocyte by FcεR1, resulting in the degranulation of the mastocytes and the excretion of preformed cellular mediators. When mastocytes are activated, they secrete ECF-A (eosinophilic chemotactic factor for anaphylaxis). Activated eosinophils excrete histaminase and arylsulfatase, enzymes which degrade histamine and SRS-A (the slow reactive substance of anaphylaxis). Eosinophils therefore limit the intensity of type 1 hypersensitivity reaction.
- GS.2.2. The main preformed mediators (deposited in mastocytes granules) are histamine (it acts on histaminergic receptors H₁, H₂, H₃), serotonin, neutrophilic chemotactic factor (NCF), eosinophilic chemotactic factor (ECF), heparin, adenosine, enzymes (tryptase).
- GS.2.3. These mediators result in pathophysiological and clinical effects: dermatological (hives, rash, pruritus, angioedema), respiratory (dyspnea, wheezing, stridor, glottic and laryngeal oedema, bronchospasm, bronchorrhea, pulmonary oedema, hypoxia), cardiovascular shock-inducing effects: arterial hypotension caused by vasodilation and hypovolemia (capillary hyperpermeability, interstitial loss of liquids), myocardial depression, myocardial ischemia, cardiac arrhythmia, digestive (nausea, emesis, diarrhea), neurologic effects (vertigo, disorientation, hallucination, convulsions)
- GS.2.4. Biphasic anaphylaxis is a clinical condition in which symptomatology repeats in approximately 6 to 8 hours (even up to 72 hours) after the primary event caused by contact with the allergen. The mechanism behind biphasic anaphylaxis is based on the excretion of a second wave of de novo mediators (arachidonic acid derivatives, PAF, etc.)
- GS.2.5. The respiratory and hemodynamic consequences are life-threatening, requiring urgent therapeutical measures: airway management, oxygen therapy, insuring ventilation, administration of bronchodilators in order to stop bronchospasm, hemodynamic resuscitation (rapid restoration of blood volume, vasoactive substance administration)
- GS.2.6. The following principles outline anaphylactic shock:
- If the anaphylaxis-causing substance is administered by IV perfusion, the administration of the respective substance must be immediately interrupted.
 - Administration of adrenaline: vasoconstrictor action (α_1 adrenergic effect), inotropic-positive (β_1 adrenergic effect), bronchodilation (β_2 adrenergic effect) pathogenetic- reduces mastocytes degranulation, through the following mechanism: increasing the activity of adenylyl cyclase (β_2 adrenergic action) increases the intracellular concentration of cAMP).
 - Supportive vital functions: oxygen therapy, tracheal intubation, mechanical ventilation, hypovolemic correction, pharmacological hemodynamic stabilization.

ChatGPT Interrogation and Output Evaluation

ChatGPT-3.5 free version was used in our study. We provide the case and ask ChatGPT to answer each individual question. We collected the answers on August 7, 2023. Since the text of the book is in Romanian, the questions were addressed to ChatGPT in Romanian and compared, later, with the GS answers

The following request was included in the ChatGPT (<https://chat.openai.com/>) window: We have the following case: "[case description presented in *Clinical case x*"]". Answer the following questions: [questions presented in *Clinical case x*]. An example of the first case is illustrated in Figure 1.

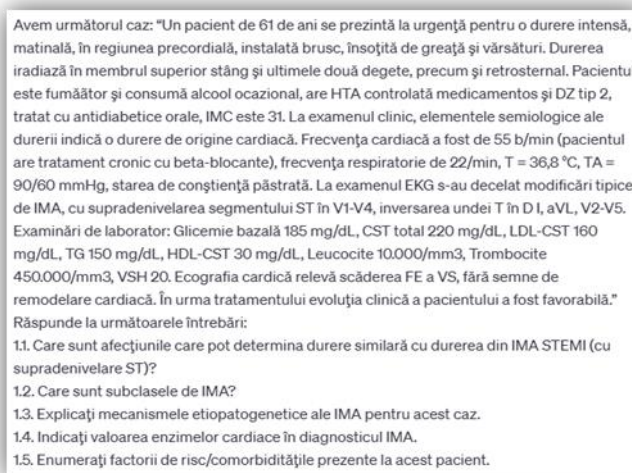


Figure 1. The ChatGPT input text: first case [in Romanian]

The numerical characteristics of the input text and gold-standard answers are given in Table 1.

Table 1. Characteristics of the evaluated cases and of gold standard responses by numerical form

Case	Case description		Gold standard responses	
	words	characters with spaces	words	characters with spaces
1	191	1208	368	2781
2	68	446	351	2938

Two individual researchers evaluated the responses given by the ChatGPT following the same criteria. The researchers classified each answer as correct, partially correct, or incorrect. Explanations were given for each classification. A consensus meeting was scheduled by a third researcher whenever there were different interpretations between researchers and the consensus result was reported.

The words in inputs and outputs were counted using Word count function (Microsoft Office 364, Word) to evaluate if the number of outputs is associated with the number of input words. GPTZero [31] (Model version: 2024-02-06) was used to test if the text sounds or not as generated by human. We evaluated the responses generated by human professionals and ChatGPT on 2 March 2024. We collected for each sample of text evaluated with GPTZero the following scores: perplexity, burstiness, readability, and simplicity. Perplexity, a metric per sentence, is a standard metric used to evaluate the performance of LLMs that measure the randomness of the text [32]. A perplexity score higher than 85 indicate a text that is more likely to be generated by LLMs [33]. Burstiness is a measurement of the variance in writing in the entire document, and a low score indicate the likelihood of LLMs generated text [33]. Readability score shows the complexity of the text and a low value indicate the need of a university degree to understood it. Simplicity score also reflects how easier to read and understand is a text and a high value indicates a higher percentage of words that belong to the 100 most common words in a specific language.

Results

Responses Given by ChatGPT

The following answers were generated by ChatGPT for the first clinical case:

ChatGPT.1.1. Conditions that can cause pain similar to that of STEMI (acute ST-segment elevation myocardial infarction) include: • Unstable or uncomplicated angina; • Acute pericarditis; • Aortic dissection; • Musculoskeletal disorders (e.g., chest pain of muscle or costosternal joints); • Pneumothorax; • Gastroesophageal reflux or esophageal conditions; • other non-cardiac causes of chest pain.

ChatGPT.1.2. The subclasses of AMI (acute myocardial infarction) include: a. AMI with ST-segment elevation (STEMI): characterized by evidence of ST-segment elevation in the EKG, signifying the complete occlusion of a large coronary artery.; b. Non-ST-segment elevation AMI (NSTEMI): ST-segment elevation is not seen on the EKG, and the coronary lesion may be partial occlusion or thrombus.

ChatGPT.1.3. The etiopathogenetic mechanisms of AMI in this case can be described as follows: • The patient has cardiovascular risk factors, such as advanced age, smoking, hypertension and type 2 diabetes mellitus, which predispose to atherosclerosis. • Atherosclerosis affects the coronary arteries, causing the formation of atheromatous plaques, which may rupture or erode. • Rupture of the atheroma plaque exposes the lipid content and blood procoagulant factors, causing the formation of a coronary thrombus. • The thrombus can completely or partially obstruct blood flow in the coronary artery, leading to ischemia and injury to the myocardial tissue downstream of the occlusion. • In the case of STEMI AMI, the thrombus causes total occlusion of the artery, leading to necrosis (cell death) in the affected area, seen on the EKG by ST-segment elevation and T-wave inversion. • The intense pain in this case is caused by myocardial ischemia and receptor activation of pain in the myocardium and adjacent tissues.

ChatGPT.1.4. The value of cardiac enzymes in the diagnosis of AMI includes: • Troponins: The most sensitive and specific biomarkers for myocardial damage. Levels rise in the first few hours of an AMI and remain elevated for days. • Creatinine kinase (CK) and its isozymes: Levels increase following myocardial injury but are not as specific as troponins. • Myoglobin: Another biomarker that increases in AMI but has less specificity than troponins.

ChatGPT.1.5. Risk factors/comorbidities present in this patient include: • Older age (61 years); • Smoking; • Occasional alcohol consumption; • Drug-controlled hypertension (HT); • Type 2 diabetes (DM) treated with oral antidiabetic agents; • Increased body mass index (BMI>31); • Disturbed lipid profile: high levels of LDL-CST (LDL cholesterol), triglycerides (TG) and low level of HDL-CST (HDL cholesterol); • Possible history of cardiovascular disease, as elevated blood pressure and DM. All these factors increase the risk of developing cardiovascular diseases, such as acute myocardial infarction.

The following answers were generated by ChatGPT for the second clinical case:

ChatGPT.2.1. Anaphylactic reactions primarily involve two types of immune cells: mast cells and basophils. These cells are sensitized to allergens and contain granules filled with inflammatory mediators. When an individual with a pre-existing allergy is exposed to the allergen (in this case, potentially a component of the Ibuprofen pill), these cells release their contents, leading to the characteristic symptoms of anaphylaxis.

ChatGPT.2.2. During an anaphylactic reaction, various mediators are released, including histamine, leukotrienes, prostaglandins, and cytokines. These mediators cause several physiological effects: • Histamine: It leads to vasodilation, increased vascular permeability, and bronchoconstriction, and is responsible for skin rash, swelling, and itching. • Leukotrienes and Prostaglandins: They contribute to smooth muscle contraction, mucus production, and further bronchoconstriction. • Cytokines: These inflammatory molecules can cause systemic symptoms and activate other immune cells.

ChatGPT.2.3. Biphasic anaphylaxis refers to a secondary phase of anaphylaxis that occurs hours after the initial reaction has been treated or subsided. This second phase can occur in a subset of patients, and its cause and mechanisms are not fully understood. It emphasizes the importance of continued monitoring and observation even after the initial treatment of an anaphylactic reaction.

ChatGPT.2.4. Vital-risk-associated clinical manifestations of anaphylactic reactions include: • Hypotension (low blood pressure); • Tachycardia (rapid heart rate); • Respiratory distress (wheezing, bronchoconstriction); • Altered consciousness.

The pathophysiology principles behind therapy involve: • Epinephrine: The first-line treatment, as it rapidly reverses vasodilation and bronchoconstriction, increasing blood pressure and improving breathing. • Fluid Resuscitation: Intravenous fluids may be administered to support blood pressure. • Bronchodilators: Medications like albuterol can help relieve bronchoconstriction. • Antihistamines and Corticosteroids: These can help counteract the effects of histamine and reduce inflammation.

ChatGPT.2.5. The principles of therapy for anaphylactic shock are to quickly reverse the life-threatening physiological changes: • Epinephrine: Administered promptly to counteract vasodilation, bronchoconstriction, and increase blood pressure. • Supportive Care: Intravenous fluids to maintain blood pressure and oxygen therapy to ensure adequate oxygen supply. • Airway Management: Assure a patent airway, and use bronchodilators as needed. • Histamine Antagonists (Antihistamines): To counteract the effects of histamine. • Corticosteroids: To reduce inflammation and prevent delayed reactions.

This patient requires immediate medical attention, including epinephrine administration, and close monitoring because of the potential for biphasic anaphylaxis. Treatment should be administered by healthcare professionals experienced in managing anaphylactic reactions.

Evaluation of Responses Given by ChatGPT

In one out of five questions, the ChatGPT response was incorrect, regardless of the investigated case. We found that the response given by ChatGPT was partially correct in most of the cases (Figure 2).

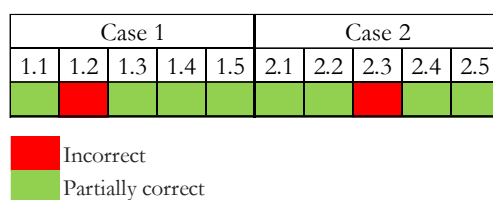


Figure 2. Analysis of the correctness of responses gave by ChatGPT (1.1-1.5 and 2.1-2.5 are the questions associated with cases)

Table 2 provides detailed explanations of the classification given in Figure 2.

The analysis of the responses showed similar quantitative scores on answers given by ChatGPT and professionals, identifying that all sets of responses are over 80% human text (Table 3).

Table 3. Quantitative score of responses by respondents

Case	Who?	Perplexity	Burstiness	Readability	Simplicity	Text is likely to be written		
						Human (%)	Mixed (%)	LLM (&)
1	ChatGPT	31.9 (medium)	55.6 (medium)	7 (low)	25.9 (low)	94	6	0
	GS	32.6 (medium)	67.2 (medium)	-19.3 (low)	23.8 (low)	91	9	0
2	ChatGPT	27.7 (low)	29 (low)	-12.7 (low)	25.6 (low)	88	12	0
	GS	31.6 (medium)	55.7 (medium)	-46.0 (low)	17.9 (low)	85	15	0

LLM = Large Language Model, GS = gold standard answer (provided by the professionals)

Discussion

Our study found that ChatGPT free version could provide partially correct answers to specific medical questions considering a specific medical case, failed in one out of five questions, and did not succeed in giving a correct answer to any question. Furthermore, the responses of ChatGPT showed similar perplexity, burstiness, readability, and simplicity metrics with human responses, metrics evaluated with GPTZero.

Table 2. Correctness of responses generated by ChatGPT for the evaluated cases

		Explanations
1.1		Pathologies missing from the ChatGPT answer: <ul style="list-style-type: none"> • Painful cardiac pathologies that can mimic cardiac ischemia: cardiomyopathies, aortic stenosis, cocaine abuse associated cardiac ischemia. • Non-painful pathologies that can cause ST segment elevation: left ventricular hypertrophy, left bundle branch block, hyperpotassemia, Brugada syndrome. • Painful pathologies that do not associate cardiac ischemia: pleuritis, costochondritis, hiatal hernia. They can be differentiated through clinical and paraclinical examinations.
1.2		ChatGPT missed the latest classification.
1.3		ChatGPT generated answer offers a broader explanation of the pathophysiological mechanisms, which is technically correct, however it does not match the expected answer, which focuses more on cellular mechanisms.
1.4		ChatGPT generated answer does not elaborate on troponin types and causes of troponin raises. On the other hand, ChatGPT generated answer gives other examples of myocardial infarction associated biomarkers such as: creatine kinase and myoglobin and gives information about each biomarker's specificity.
1.5		ChatGPT generated answer includes all the risk factors listed in the gold-standard answer and add age, alcohol consumption and possibility of patients' history of hypertension and diabetes mellitus.
2.1		The ChatGPT-generated answer offers a broader explanation of the pathophysiological mechanisms, which, while still correct, differs from the original answer and do not describe the mechanism behind type 1 hypersensitivity.
2.2		The ChatGPT-generated answer does not include serotonin, heparin, adenosine, and enzymes such as tryptase in the preformed mediators, but it adds leukotrienes and prostaglandins. The gold standard answer also offers more possible effects of the mediators.
2.3		The gold standard answer provides the cause and mechanism behind biphasic anaphylaxis and a more exact hour interval in which the secondary phase of anaphylaxis may occur.
2.4		The ChatGPT-generated answer provides a more complete and detailed account of vital-risk-associated clinical manifestations compared to the original answer. When describing the physiopathology principles behind therapy, the ChatGPT fails to mention oxygen therapy and airway management.
2.5		The ChatGPT-generated response does not mention that IV administration of anaphylaxis reaction-causing substances should be immediately interrupted. It also does not explain the mechanism behind epinephrine action. On the other hand, the ChatGPT-generated answer gives other examples of therapeutic methods.

Our study showed the capacity of ChatGPT to address specific questions on specified scenarios, capturing in most of the cases at least parts of correct answers (Figure 2, Table 2) in a scenario with generated responses in Romanian language considering that scientific medical information is available in English language. Comparing the number of words and characters with spaces on the ChatGPT and GS responses, it can be seen that are similar, with slightly higher numbers in case of ChatGPT, with a difference of 16 (case 2) and respectively 32 (case 1) words in the favor of ChatGPT. The difference in terms of characters with spaces is less than 3% higher characters in ChatGPT responses than in professionals' responses. When looking at the number of words and characters, ChatGPT showed the capacity to generate responses similar to professionals but faster. One of ChatGPT feature is the production of formally structured text using eloquent vocabulary [37], also observed in our study (Table 3). The readability and simplicity of the responses generated by ChatGPT is similar to those written by the professionals (Table 3), showing an adaptation to medical vocabulary and destination of text to medical literate readers. In the first investigated case, the perplexity and burstiness show the same level of closeness to large language models between ChatGPT and human generated responses but with a different class for the second case (Table 3). Our results indicate the incapacity of GPTZero to identify LLMs generated text, with classification as being written by human, with higher percentages obtained by ChatGPT generated text (Table 3). Our results align with what is already known from scientific literature regarding the performances in scientific writing, specifically that ChatGPT can assist in some extent in scientific writing, but the intervention of humans is a must [34,35].

Even though ChatGPT was not developed to assist physicians or patients towards healthcare, it is capable to generate text on a wide range of topic, including medical information [36]. ChatGPT have several applications that could fit into medicine, especially in education (implications in generation of virtual patients, quizzes, skills assessment etc.) [37,38], patient's conversational agent [39], patient monitoring and follow-up [40], or a virtual

assistant health literacy training [41]. Shen et al. [42] demonstrated the superiority of ChatGPT responses to fact, policies and diagnosis questions related with otolaryngology-head and neck surgery, with lower readability ($p < 0.001$) and similar understandability ($p = 0.17$) (Patient Education Materials Assessment Tool). The responses provided by ChatGPT to patient questions related to symptom-based diagnoses outperformed the web search [42].

It is worth to mention that sometimes responses given by ChatGPT are not accurate, phenomenon known as "artificial hallucination or confabulation" [43,44,45]. We did not encounter such deviation in our study (Table 2), maybe because we ask ChatGPT to produce responses in a strict and clear context and guidance. Furthermore, we did not ask ChatGPT to provide references to support the responses, because its limitations in this regard have been previously reported in the scientific literature [43,46]. However, ChatGPT could become a potential source of medical information since it can provide to some extent valid answers even if the responses are not as detailed as expert's answers. It should be noted that low readability and simplicity could be a benefit both for students and the general population, showing a high potential for understanding the concepts. ChatGPT is a 24/24 and 7/7 available resource, a chameleon that adapts to several languages and ready-to-use virtual assistant with potential in learning both for medical professionals and general population. Furthermore, ChatGPT could assist medical teaching staff to create tests, case scenarios, training materials or to make summaries [47]. Despite the inability of ChatGPT to teach integration of contextual information or sensory data or other nonverbal cues [48], it is expected that tools such ChatGPT to find their place also in medical education. However, international, national, and institutional guidelines in both teaching and learning by respecting integrity, ethics, and professionalism must be put in place and implemented [49,50].

Our study has several strengths that deserve to be underlined. Evaluation of ChatGPT performance is a must to know what are its abilities and performances in specific domains. Even that it was not constructed as an assistant in learning and teaching medicine, the tool exists so it is expected that at least some students and teaching staff to used it [51,52]. As a consequence, a new area of research is open and to date a limited number of scientific references were reported regarding the use and performances of ChatGPT in Romanian language [53,54]. The scientific literature lacks evidence in using ChatGPT as an assistant in learning medicine, so our study is, to be best of our knowledge, the first full article investigating its abilities with input text in Romanian language. Another strength is represented by quantitative metrics used to evaluate the performance of the responses received from ChatGPT. Even though metrics allow an objective evaluation, these metrics are still not evidence and indicate only the possibility and need to be produced by humans.

Our study has several limitations that must be emphasized. First, our study is limited to two pathophysiological cases using the free version of ChatGPT. In consequence, the generalizability of our findings is limited. We expected a similar number of inaccuracies when the number of examples increased but this aspect needs further evaluation. The performances could be different when cases with different complexity are the input or when the paid version of ChatGPT is used. Furthermore, differences between free (ChatGPT-3.5) and paid version (ChatGPT-4) should also be evaluated. Second, the language of input and output was Romanian, and it is worth to investigate if performances of the outputs are different when the input text is in English language. Third, we acknowledge that the ChatGPT is not specifically created as a learning assistant for medicine and a comparative study with a dedicated LLM tool such PubMed Buddy [55] or Med-PaLM 2 [56], a tool that provide access to full-text scientific articles, would provide appropriate evidence of its performances.

Overall, ChatGPT performed satisfactorily, being incorrect in only one question out of five, and would be of interest to compare its performance with medical student's performances. We expect ChatGPT and other large language models to make progress and change also medical training and teaching. Besides opportunities opened by ChatGPT, its threats related to medical training must be acknowledged. As an emergencies technology available for free, we must train the users about the use and misuses of ChatGPT in academic environment, acknowledging opportunities, strengths, weaknesses, and threats and their reflections in academic integrity. Our study showed that ChatGPT can be a valuable instrument for answering specific questions related to an explicit medical case, but experts' human intervention is needed because correctness is not full.

List of Abbreviations: AI- artificial intelligence; AMI- Acute Myocardial Infarction; BMI- Body Mass Index; BP- blood pressure; CK- Creatinine kinase; CST- cholesterol; DM- Diabetes Mellitus; ECF- eosinophilic chemotactic factor; EKG-

electrocardiogram; ESR- erythrocytes sedimentation rate; HT- hypertension; LVEF- Left Ventricular Ejection Fraction; NCF- neutrophilic chemotactic factor; NSTEMI- not acute ST-segment elevation myocardial infarction; SRS-A- slow reactive substance of anaphylaxis; STEMI- acute ST-segment elevation myocardial infarction; T- body temperature; TG- triglycerides.

Author Contributions: SDB defined the research's aim and the experiment's design. SDB, AEB and IS participated in the design of the study. AIB, BB, AEB and IS carried out the experiments. AIB and BB draft the manuscript. SDB, AEB and IS reviewed the manuscript. SDB coordinated the manuscript. AEB supervised the project. All authors read and approved the final manuscript.

Funding: This research received no funding.

Ethics Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We used ChatGPT (3.5) to generate the answers and GPTZero (2024-02-06) to evaluate the responses in terms of perplexity, burstiness, readability and simplicity.

Conflict of Interest: The authors declare no conflict of interest.

References

1. Karako K, Song P, Chen Y, Tang W. New possibilities for medical support systems utilizing artificial intelligence (AI) and data platforms. *Biosci Trends*. 2023;17(3):186-189. doi: 10.5582/bst.2023.01138.
2. Ministry of Health, Labour and Welfare. Data-based Health Management Initiatives <https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000148743.html> (Accessed March 19, 2024).
3. Australian Digital Health Agency, My Health Record, Australian Digital Health Agency <https://www.digitalhealth.gov.au/initiatives-and-programs/my-health-record> (Accessed March 19, 2024)
4. NHS Digital. Home - NHS Digital Available from: <https://digital.nhs.uk/> (Accessed March 19, 2024)
5. Dwivedi YK, Kshetri N, Hughes L, Slade EL, Jeyaraj A, Kar AK, et al. Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 2023;71:102642. doi:10.1016/j.ijinfomgt.2023.102642
6. Welsh S. [internet] 2023 [accessed on March 23, 2024] ChatGPT and the Future of the Professions. Quillette. Available from: <https://quillette.com/2023/01/26/chatgpt-and-the-future-of-the-professions/?ref=quillette-newsletter>
7. Funelas R. ChatGPT Language Capabilities: The Breakdown [31 January 2024] (accessed 25 March 2024) Available from: <https://www.tomedes.com/translator-hub/chatgpt-language-capabilities>
8. Shrestha N, Shen Z, Zaidat B, Duey AH, Tang JE, Ahmed W, Hoang T, Restrepo Mejia M, Rajjoub R, Markowitz JS, Kim JS, Cho SK Performance of ChatGPT on NASS Clinical Guidelines for the Diagnosis and Treatment of Low Back Pain: A Comparison Study. *Spine (Phila Pa 1976)*. 2024. doi: 10.1097/BRS.0000000000004915.
9. Patrinos GP, Sarhangi N, Sarrami B, Khodayari N, Larijani B, Hasanzad, Using ChatGPT to predict the future of personalized medicine. *M.Pharmacogenomics J*. 2023;23(6):178-184. doi: 10.1038/s41397-023-00316-9
10. Wójcik S, Rulkiewicz A, Pruszczyk P, Lisik W, Poboży M, Domienik-Karłowicz J. *Cardiol J*. 2023;30(6):1018-25. doi: 10.5603/cj.97515
11. Christiano PF, Leike J, Brown T, Martic M, Legg S, Amodei D. Deep reinforcement learning from human preferences. In: Guyon I, Von Luxburg U, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R. *Advances in Neural Information Processing Systems* 30. 2017, p. 4299- 4307.
12. Baidoo-Anu D, Owusu Ansah L. Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning. *Journal of AI*. 2023;7(1):52-62
13. Zhuo TY, Huang Y, Chen C, Xing Z. Exploring Ai Ethics of Chatgpt: A Diagnostic Analysis. arXiv. 2023. [accessed 23 March 2024] Available from: <https://ar5iv.labs.arxiv.org/html/2301.12867>
14. Lecler A, Duron L, Soyer P. Revolutionizing radiology with GPT-based models: Current applications, future possibilities and limitations of ChatGPT. *Diagnostic and Interventional Imaging* 2023;104(6):269-274. doi:10.1016/j.diii.2023.02.003.

15. Gozalo-Brizuela R, Garrido-Merchan EC. ChatGPT is not all you need. a state of the art review of large generative AI models. 2023 [accessed 23 March 2024]. Available from: <https://arxiv.org/pdf/2301.04655.pdf>
16. Ray PP. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems* 2023;3:121-154. doi: 10.1016/j.iotcps.2023.04.003.
17. Sison AJG, Daza MT, Gozalo-Brizuela R, Garrido-Merchán EC. ChatGPT: More Than a “Weapon of Mass Deception” Ethical Challenges and Responses from the Human-Centered Artificial Intelligence (HCAI) Perspective. *International Journal of Human-Computer Interaction* 2023. doi: 10.1080/10447318.2023.2225931.
18. O’Connor S. Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse? *Nurse Education in Practice* 2023;66:103537. doi:10.1016/j.nepr.2022.103537.
19. Mijwil MM, Hiran KK, Doshi R, Dadhich M, Al-Mistarehi A-H, Bala I. ChatGPT and the future of academic integrity in the artificial intelligence era: A new frontier. *Al-Salam Journal for Engineering and Technology* 2023;2(2):116-127. doi:10.55145/ajest.2023.02.02.015.
20. Johnson D, Goodman R, Patrinely J, Stone C, Zimmerman E, Donald R, et al. Assessing the Accuracy and Reliability of AI-Generated Medical Responses: An Evaluation of the Chat-GPT Model. *Res Sq.* 2023;rs.3.rs-2566942. doi: 10.21203/rs.3.rs-2566942/v1
21. Ismail AMA. Chat GPT in Tailoring Individualized Lifestyle-Modification Programs in Metabolic Syndrome: Potentials and Difficulties? *Ann Biomed Eng.* 2023;51(12):2634-2635. doi: 10.1007/s10439-023-03279-x
22. Gutiérrez-Cirlos C, Bermúdez-González JL, Carrillo-Pérez DL, Hidrogo-Montemayor I, Martínez-González A, Carrillo-Esper R, Sánchez-Mendiola M. Medicine and the metaverse: current applications and future. *Gac Med Mex.* 2023;159(4):280-286. doi: 10.24875/GMM.M23000795
23. Klang E, Portugez S, Gross R, Kassif Lerner R, Brenner A, Gilboa M, et al. Advantages and pitfalls in utilizing artificial intelligence for crafting medical examinations: a medical education pilot study with GPT-4 *BMC Med Educ* 2023;23(1):772. doi: 10.1186/s12909-023-04752-w
24. Zybaczynska J, Norris M, Modi S, Brennan J, Jhaveri P, Craig TJ, Al-Shaikhly T. Artificial Intelligence-Generated Scientific Literature: A Critical Appraisal. *J Allergy Clin Immunol Pract.* 2024;12(1):106-110. doi: 10.1016/j.jaip.2023.10.010
25. Grewal H, Dhillon G, Monga V, Sharma P, Buddhavarapu VS, Sidhu G, et al. Radiology Gets Chatty: The ChatGPT Saga Unfolds. *Cureus.* 2023;15(6):e40135. doi: 10.7759/cureus.40135
26. Ali K, Barhom N, Tamimi F, Duggal M. ChatGPT-A double-edged sword for healthcare education? Implications for assessments of dental students. *Eur J Dent Educ.* 2024;28(1):206-211. doi: 10.1111/eje.12937
27. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, Chartash D. How Does ChatGPT Perform on the United States Medical Licensing Examination? The Implications of Large Language Models for Medical Education and Knowledge Assessment. *JMIR Med Educ.* 2023;9:e45312. doi: 10.2196/45312
28. Bartoli A, May AT, Al-Awadhi A, Schaller K. Probing artificial intelligence in neurosurgical training: ChatGPT takes a neurosurgical residents written exam. *Brain Spine.* 2023;4:102715. doi: 10.1016/j.bas.2023.102715. eCollection 2024.
29. Bulboacă AE. [Fiziopatologie aplicata] [in Romanian]. Vol II, Ed Risoprint, Cluj-Napoca, Romania, 2023.
30. Domienik-Karłowicz J, Kupczyńska K, Michalski B, Kaplon-Cieślicka A, Darocha S, Dobrowolski P, et al. Fourth universal definition of myocardial infarction. Selected messages from the European Society of Cardiology document and lessons learned from the new guidelines on ST-segment elevation myocardial infarction and non-ST-segment elevation-acute coronary syndrome. *Cardiol J.* 2021;28(2):195-201. doi: 10.5603/CJ.a2021.0036.
31. Tian E, Cui A. GPTZero: Towards detection of AI-generated text using zero-shot and supervised methods. *GPTZero.* 2023 Available from: <https://gptzero.me> (accessed August 20, 2023)
32. Miaschi A, Brunato D, Dell’Orletta F, Venturi G. What Makes My Model Perplexed? A Linguistic Investigation on Neural Language Models Perplexity. *Proceedings of Deep Learning Inside Out (DeeLIO): The 2nd Workshop on Knowledge Extraction and Integration for Deep Learning Architectures.* 2021, pp. 40-47. doi:10.18653/v1/2021.deelio-1.5

33. How do I interpret burstiness or perplexity? [Internet] [accessed 20 March 2024] Available from: <https://support.gptzero.me/hc/en-us/articles/15130070230551-How-do-I-interpret-burstiness-or-perplexity>
34. Else H. Abstracts written by ChatGPT fool scientists. *Nature* 2023;613:423. doi: 10.1038/d41586-023-00056-7
35. Blanco-González A, Cabezón A, Seco-González A, Conde-Torres D, Antelo-Riveiro P, Piñeiro Á, Garcia-Fandino R. The Role of AI in Drug Discovery: Challenges, Opportunities, and Strategies. *Pharmaceuticals*. 2023;16(6):891. doi:10.3390/ph16060891
36. Radford A, Narasimhan K, Salimans T, Sutskever I. Improving language understanding by generative pre-training. [Internet] 2018 [accessed 2 March 2024]. Available from: <https://www.mikecaptain.com/resources/pdf/GPT-1.pdf>
37. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Front Artif Intell*. 2023;6:1169595. doi: 10.3389/frai.2023.1169595.
38. Han J-W, Park J, Lee H. Analysis of the effect of an artificial intelligence chatbot educational program on non-face-to-face classes: a quasi-experimental study. *BMC Med. Educ*. 2022;22:830. 10.1186/s12909-022-03898-3.
39. Issom D-Z, Hardy-Dessources M-D, Romana M, Hartvigsen G, Lovis C. Toward a conversational agent to support the self-management of adults and young adults with sickle cell disease: usability and usefulness study. *Front. Digit. Health* 2021;3:600333. 10.3389/fdgh.2021.600333
40. Eysenbach G. The Role of ChatGPT, Generative Language Models, and Artificial Intelligence in Medical Education: A Conversation With ChatGPT and a Call for Papers. *JMIR Med Educ*. 2023;9:e46885. doi: 10.2196/46885.
41. Mokmin NAM, Ibrahim NA. The evaluation of chatbot as a tool for health literacy education among undergraduate students. *Educ Inf Technol (Dordr)* 2021;26:6033–6049. doi: 10.1007/s10639-021-10542-y
42. Shen SA, Perez-Heydrich CA, Xie DX, Nellis JC. ChatGPT vs. web search for patient questions: what does ChatGPT do better? *Eur Arch Otorhinolaryngol*. 2024. doi: 10.1007/s00405-024-08524-0.
43. Athaluri SA, Manthena SV, Kesapragada VSRKM, Yarlagadda V, Dave T, Duddumpudi RTS. Exploring the Boundaries of Reality: Investigating the Phenomenon of Artificial Intelligence Hallucination in Scientific Writing Through ChatGPT References. *Cureus*. 2023;15(4):e37432. doi: 10.7759/cureus.37432
44. Jamaluddin J, Gaffar NA, Din NSS. Hallucination: A key challenge to Artificial Intelligence-Generated writing. *Malays Fam Physician*. 2023;18:68. doi: 10.51866/lte.527.
45. Conroy G. Scientists used ChatGPT to generate an entire paper from scratch - but is it any good? *Nature*. 2023;619(7970):443-444. doi: 10.1038/d41586-023-02218-z.
46. Buholayka M, Zouabi R, Tadinada A. The Readiness of ChatGPT to Write Scientific Case Reports Independently: A Comparative Evaluation Between Human and Artificial Intelligence. *Cureus*. 2023;15(5):e39386. doi: 10.7759/cureus.39386.
47. Jeyaraman M, K SP, Jeyaraman N, Nallakumarasamy A, Yadav S, Bondili SK. ChatGPT in Medical Education and Research: A Boon or a Bane? *Cureus*. 2023;15(8):e44316. doi: 10.7759/cureus.44316
48. Safranek CW, Sidamon-Eristoff AE, Gilson A, Chartash D. The Role of Large Language Models in Medical Education: Applications and Implications. *JMIR Med Educ*. 2023;9:e50945. doi: 10.2196/50945.
49. van de Ridder JMM, Shoja MM, Rajput V. Finding the Place of ChatGPT in Medical Education. *Academic Medicine* 2023;98(8):867. doi:10.1097/ACM.0000000000005254.
50. Wong RS, Ming LC, Raja Ali RA. The Intersection of ChatGPT, Clinical Medicine, and Medical Education. *JMIR Med Educ*. 2023;9:e47274. doi: 10.2196/47274.
51. Cross J, Robinson R, Devaraju S, Vaughans A, Hood R, Kayalackakom T, et al. Transforming Medical Education: Assessing the Integration of ChatGPT Into Faculty Workflows at a Caribbean Medical School. *Cureus* 2023;15(7):e41399. doi:10.7759/cureus.41399.
52. Hosseini M, Gao CA, Liebovitz DM, Carvalho AM, Ahmad FS, Luo Y, et al. An exploratory survey about using ChatGPT in education, healthcare, and research. *PLoS One*. 2023;18(10):e0292216. doi: 10.1371/journal.pone.0292216.
53. Bulboacă AI, Bulboacă AE, Stănescu I, Bolboacă SD. ChatGPT Performances in Explanation of Physiopathology by Example. *Appl Med Inform*. 2023;45(Suppl. S1):S33.
54. Bolboacă SD. ChatGPT Assistance in Academic Assignments by Example. *Appl Med Inform*. 2023;45(Suppl. S1):S31.

55. Hatfield S. PubMed Buddy [internet] [accessed 25 March 2024] Available from: <https://chat.openai.com/g/g-5KWeLiLYG-pubmed-buddy>
56. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. *Nature* 2023;620:172-180. doi:10.1038/s41586-023-06291-2