



## Assessment of ChatGPT as A Potential Tool In Solving Case Scenarios And Knowledge-Based Questions In Pathology

<sup>1</sup>Dr. Isha Aggarwal, <sup>2</sup>Dr. Shweta Rana, <sup>3</sup>Dr. Amrita Kulhria, <sup>4</sup>Dr. Preeti Patni, <sup>5</sup>Dr. Mayank Aggarwal

M.D (Pathology), <sup>1,3</sup>Assistant Professor, <sup>2</sup>Associate Professor,

<sup>4</sup>Professor & Head, <sup>5</sup>Senior Resident,

Department Of Pathology,

Shri Atal Bihari Vajpayee Government Medical College, Chhainsa Faridabad, Haryana

**\*Corresponding Author:**

**Dr. Shweta Rana**

Associate Professor, Department Of Pathology,

Shri Atal Bihari Vajpayee Government Medical College, Chhainsa Faridabad, Haryana

Type of Publication: Original Research Paper

Conflicts of Interest: Nil

### Abstract

**Background:** ChatGPT (Chat Generative Pre-Trained Transformer) is an advanced conversational Artificial Intelligence tool that can support educational initiatives. The aim of this study is to provide a comprehensive and practical assessment of ChatGPT for the application of AI in medical education.

**Material and Methods:** A total of 100 questions from 36 topics of Pathology according to CBME (Competency based medical education) curriculum were selected by three faculty members. These 100 questions comprised 50 open-ended questions which included problem-based questions and long answer questions and 50 multiple choice questions (MCQs) with single answer. The response generated by ChatGPT were evaluated. Using a score of 0-5, all 100 questions were evaluated.

**Result:** Out of 50 MCQs, answers of 39 (78%) were found correct along with appropriate explanations. The mean score was 3.9 with standard deviation of 2.09. The result was statistically significant with a mean hypothetical value of 2.5 but was insignificant with mean hypothetical value of 4.0. In the case of 50 long answer questions, the overall mean score was 4.36 with standard deviation of 0.62, which was found to be statistically significant.

**Conclusion:** ChatGPT can serve as virtual tutor, mentor and guide offering explanations of complex medical concepts requiring higher order reasoning. However, ChatGPT seems partially effective in providing relevant details specific for pathology in terms of laboratory diagnosis. Further improvement, review and validation is required for its potential usage.

**Keywords:** ChatGPT, Artificial intelligence, Competency based medical education, Pathology, Multiple choice questions

### Introduction

The term artificial intelligence (AI) was first coined by John McCarthy in 1955 [1]. But in recent years, with the advent of new algorithms such as machine learning (ML) and its subsets, AI has evolved significantly. Artificial intelligence (AI) refers to the simulation of human intelligence processes by machines, particularly computer systems [2]. One of the key components of AI is the Natural Language

Processing (NLP) that focuses on enabling computers to understand, interpret and generate human language in a way that is both meaningful and contextually relevant. These days, AI is transforming education through adaptive learning platforms, intelligent tutoring systems, automated grading and personalized learning experience tailored to individual student's need [3].

ChatGPT (Chat Generative Pre-Trained Transformer) is one such advanced conversational AI developed by Open AI that debuted to the public in November 2022 [4]. It can support educational initiatives by providing explanations, answering questions and offering personalized learning experience to students. Medical education includes various levels, such as undergraduate, postgraduate, specialty, and subspecialty education, also known as continued medical education (CME) [5]. Pathology is a crucial subject in MBBS curriculum that deals with the study of diseases and their effects on the structure and function of the body. The new CBME 2019 curriculum of Pathology comprises 36 topics that have 182 competencies [6]. Understanding Pathology is important as it provides medical students with essential knowledge and skills for understanding disease processes and making accurate diagnosis. The use of ChatGPT as an automated question answering system in the field of medical education could aid in the development of effective educational tools that can help students to learn and understand the subject better. To achieve educational objectives, it is important to comprehensively understand the advantages and disadvantages of incorporating AI in medical education, recognizing favorable and unfavorable factors that go hand in hand. The aim of this study is to provide a comprehensive and practical review of the benefits, challenges, and solutions for the application of AI in medical education.

Hence in light of this, we aimed to evaluate the use of ChatGPT in solving long answer questions (open ended) like clinical scenarios or problem-based questions, knowledge-based questions and closed questions comprising of multiple-choice questions in the subject of Pathology, so that ChatGPT can serve as a versatile and interactive study companion for MBBS students in supporting their learning and professional development.

## Material And Methods

This observational study was conducted in the Department of Pathology at Shri Atal Bihari Vajpayee Government Medical College, Chhainsa, Faridabad using the ChatGPT 3.5 version which is an open-source free online tool for users after registration. A total of 100 questions from 36 topics of Pathology according to CBME curriculum were selected by three faculty members. These 100 questions comprised 50

open ended questions which included problem-based questions and long questions about definition, pathogenesis, diagnosis, differential diagnosis and laboratory diagnosis of diseases and 50 closed ended questions which included multiple choice questions with single answer. These questions were then entered into ChatGPT by a single user after free registration and the first response generated by ChatGPT was stored in the computer for further analysis. The responses generated by ChatGPT were evaluated by 2 faculty members of Pathology independently. Both faculty members were blinded towards evaluation by each other. Using a score of 0-5, all 100 questions were scored, where score 0 meant incorrect, score of 2.5 meant 50% of written answer was correct, score of 4 meant 80% of answer was correct and score 5 meant 100% of answer was correct. The average of two scores given by the two faculty members was taken as the final score.

## Results

Among 50 multiple choice questions, answers of 39 MCQs (78%) were found correct along with appropriate explanations. Out of 11 incorrect answers, in 2 MCQs answer selected was wrong but in explanation part, correct answer was explained. In the remaining 9 incorrect answers, the selected answers as well as the explanations were incorrect. Thus, the mean score in the case of MCQs was 3.9 with standard deviation of 2.09. When the data analysis was done by unpaired t test with a hypothetical value of 2.5, 4.0 and 5.0 using a Graph Pad Prism software p value was <0.0001, <0.7366 and =0.0005 respectively. Hence, the results for MCQs were found to be statistically significant with a mean hypothetical value of 2.5 but were found insignificant with mean hypothetical value of 4.0 and extremely statistically significant with mean hypothetical value of 5.0.

Out of 50 long answer questions, answers generated for 45 questions (90%) scored 4 or more out of 5 marks. In the remaining 5 questions, in one question the diagnosis was incorrect, in another the diagnosis was correct but explanation was wrong and in rest of the three questions pathogenesis was incompletely explained and in laboratory diagnosis, histopathological features were not explained at all. The histopathological aspect is important in pathological diagnosis which was not properly explained in almost all the long answer questions. In

the case of 50 long answer questions, the overall mean score was 4.36, median 4.5, mode 5 and standard deviation is 0.62.

When the data analysis was done by unpaired t test with a hypothetical value of 2.5 and 4.0 using a Graph Pad Prism software p value <0.0001 and <0.0002 was found extremely statistically significant in case of long answer questions.

## Discussion

Research has consistently demonstrated that medical students make extensive use of point-of-care tools and mobile medical applications to support their learning, citing knowledge augmentation, reliability, and efficiency. While a dedicated mobile application is not currently available, ChatGPT can be accessed through web browsers on mobile devices, allowing users to utilize the tool on-demand [7]. One of the key competencies expected of a medical graduate is the ability to generate a reasonable diagnosis for a given clinical presentation. Multiple-choice examinations represent the most common examination format in medical school and is now mandatory for undergraduate as per the new Assessment Module for Undergraduate Medical Education Training Program, 2019 issued by Medical council of India [8]. A study by Lee suggests that ChatGPT can be utilized in this context as a “virtual teaching assistant” to offer insight into each question and provide feedback on questions and clarifying concepts the student may find particularly challenging [9]. Several other studies found that ChatGPT can provide answers to questions in the subjects of Microbiology [10], Biochemistry [11] and Pharmacology [12]. A study by Sinha et al [13] in the subject of Pathology also showed that ChatGPT can solve higher order reasoning questions and the answers had a rational level of accuracy. Hence ChatGPT can be used as a valuable tool in medical education for offering several benefits. ChatGPT can serve as a virtual tutor or mentor to the medical students in explaining complex medical concepts, answering different types of questions and in explaining reasons for correct answers in case of multiple-choice questions. Students can interact with ChatGPT in real time, receiving immediate answers and explanations for answers. In this study, out of 50 long questions, the answer to only 1 question was incorrect (Figure 1) and in another one, diagnosis was correct, but explanation was incorrect (Figure 2) while

90% of answers were correctly generated by Chat GPT. The answers generated to long open-ended questions were mostly generalized and covered medicinal, surgical and radiological aspects, thus providing a multidisciplinary approach towards solving knowledge-based questions. It was found that ChatGPT can be utilized as a tool to broaden their differentials and trigger self-directed reading on diagnoses unfamiliar to the student. In the clinical environment, students can enter pertinent elements of the patient's history to generate a provisional differential diagnosis. However, it was conspicuous that in most of the questions where in laboratory diagnosis was discussed and biopsy was mentioned; histopathological features of lesions were not explained at all. ChatGPT is an effective general resource where knowledge-based questions are clearly elaborated in the prompt. For students who speak English as a second language, ChatGPT can assist in framing answers for questions and in understanding difficult medical terminology. However as depicted in our study, among the answers to knowledge-based questions, pathogenesis was also not explained properly in some of the questions where it was asked. In this study, out of 50 multiple choice questions which were solved through Chat GPT, answers generated to 39 questions were found correct along with correct explanations. Out of 11 incorrect answers, in 2 questions it was found that explanation was correct, yet correct option was not selected which is beyond comprehension (Figure 3). The 9 answers found incorrect (Figure 4) could be due to limitation of information, non-availability of recent advances and inability to recognize and understand patterns, meaning and context of information as noticed similarly by Sinha et al [13]. Self-directed learning refers to a process where individuals take the initiative and are responsible for their learning outcome, rather than relying on external authorities to dictate their educational path. According to the new CBME curriculum introduced by MCI in 2019, self-directed learning is an important part of curriculum [14]. ChatGPT has the potential to significantly enhance self-directed learning by interactive learning experience through virtual tutor, language processing tools that provide explanations and answers to all types of questions. Thus, artificial intelligence has the potential to empower learners to achieve their educational goals more effectively than ever before.

Overall, ChatGPT can play a valuable role in supplementing traditional educational methods of learning. However, it is essential to recognize the limitations of artificial intelligence as analyzed in our study where in case of MCQ, 22% of answers generated were incorrect. Although ChatGPT has the potential to revolutionize medical education, it is important to note that AI technology cannot replace human expertise and judgement. Medical students still need to be trained by experienced practitioners to develop their clinical skills and gain a deep understanding of the field. ChatGPT cannot replace the hands-on experience and mentorship that is essential for effective medical training. Thus, it is important to ensure that use of AI in education should be accompanied by adequate human support for validation to ensure accuracy and reliability and supervision to maximize its benefits effectively. AI cannot create original, innovative ideas without human input [15]. Faculty should advocate responsible and ethical use of AI and guide students through its benefits and pitfalls in courses and curriculum.

## Conclusion

The use of ChatGPT in Pathology offers numerous advantages and learning support to students. ChatGPT can serve as virtual tutor, mentor and guide in medical institutes with limited resources. Thus, by offering real time guidance through explanations of complex medical concepts requiring higher order reasoning, it can enhance the self- directed learning process. However, it is crucial to recognize that AI has its own limitations and its use in medical education requires appropriate supervision, validation for its integration in the existing educational framework.

## References:

1. McCarthy J, Minsky M, Rochester N, Shannon C. A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. *AI Magazine*. 2006; 27 (4):12-14. Available from:<doi.org/10.1609/aimag.v27i4.1904>
2. Wang F, Preininger A. AI in health: state of the art, challenges, and future directions. In: *Yearbook of medical informatics*. 2019; 28(01):16–26. Available from:< doi: 10.1055/s-0039-1677908>
3. Niazi MK, Parwani AV, Gurcan MN: Digital pathology and artificial intelligence. *Lancet Oncol*. 2019; 20(5):253-61. Available from:<doi: 10.1016/S1470-2045(19)30154-8>
4. Introducing ChatGPT. OpenAI. Accessed date March 23, 2023. Available from: URL ://openai.com/blog/chatgpt.
5. Zarei M, Mamaghani HE, Abbasi A, Hosseini MS. Application of artificial intelligence in medical education: A review of benefits, challenges, and solutions. *Medicina Clínica Práctica*. 2024; 7(2). Available from:< doi:10.1016/j.mcpsp.2023.100422>
6. <https://www.nmc.org.in/information-desk/for-colleges/ug-curriculum>
7. Tsang, R. Practical Applications of ChatGPT in Undergraduate Medical Education. *Journal of Medical Education and Curricular Development*. 2023; 10: 1-3. Available from:< doi:10.1177/23821205231178449>
8. Assessment Module for Undergraduate Medical Education Training Program. Medical Council of India, 2019.1-29.
9. Lee H. The rise of ChatGPT: exploring its potential in medical education. *Anat Sci Educ*. 2023. Available from:<doi: 10.1002/ase.2270>
10. Das D, Kumar N, Longjam L, et al. Assessing the Capability of ChatGPT in Answering First- and Second-Order Knowledge Questions on Microbiology as per Competency-Based Medical Education Curriculum. *Cureus*. 2023; 15(3). Available from:<doi: 10.7759/cureus.36034>
11. Ghosh A and Bir A. Evaluating ChatGPT's Ability to Solve Higher-Order Questions on the Competency-Based Medical Education Curriculum in Medical Biochemistry. *Cureus*. 2023;15(4). Available from:< doi: 10.7759/cureus.37023>
12. Juhi A, Pipil N, Santra S, et al. The Capability of ChatGPT in Predicting and Explaining Common Drug-Drug Interactions. *Cureus*. 2023; 15(3). Available from:< doi:10.7759/cureus.36272>
13. Sinha R K, Deb Roy A, Kumar N, et al. Applicability of ChatGPT in Assisting to Solve Higher Order in Pathology. *Cureus*. 2023; 15(2). Available from:< doi:10.7759/cureus.35237>
14. Competency-Based Undergraduate Curriculum for the Indian Medical Graduate. New Delhi, India: Council of India (MCI); 2018.
15. Karn A, Priyadarshi A, Roy AD: A review on digitalization of healthcare with SWOC analysis of digital pathology in the backdrop of COVID-19.



Global J Res Analysis. 2022, 11:1-2.  
10.36106/gjra/8400164.

## Figures

**Figure 1: Screenshot showing incorrectly diagnosed long answer question**

ChatGPT  
Based on the clinical presentation and laboratory investigations:

- 1) The likely diagnosis is **"acute myeloid leukemia (AML)"**. The patient's symptoms of headache and fever, along with laboratory findings of elevated total leukocyte count (TLC), thrombocytosis, and abnormal myeloid cell morphology (presence of metamyelocytes and myelocytes) in the differential leukocyte count (DLC), are indicative of AML, a type of blood cancer characterized by the rapid growth of abnormal myeloid cells in the bone marrow.
- 2) The next best investigation to be advised to the patient is **"bone marrow aspiration and biopsy"**. This procedure involves collecting a sample of bone marrow cells from the hip bone (usually) and examining them under a microscope to confirm the diagnosis of AML, determine the subtype, and assess the extent of involvement.
- 3) Two differential diagnoses of AML include:
  - a. **"Acute lymphoblastic leukemia (ALL)"**: Another type of blood cancer characterized by the rapid growth of abnormal lymphoblasts (immature lymphocytes) in the bone marrow. ALL is more common in children but can also occur in adults.
  - b. **"Chronic myeloid leukemia (CML)"**: A type of blood cancer characterized by the excessive proliferation of mature myeloid cells, particularly granulocytes, in the bone marrow. Unlike AML, CML typically progresses slowly and may initially present with nonspecific symptoms or be asymptomatic.

**Figure 2: Screenshot showing correct diagnosis with incorrect explanation in one long answer question.**

User  
A 28-year-old woman belongs to Bengali family has Hemoglobin of 10 g/dl. She has no history of any unusual bleeding. Other findings are MCV - 55fL, MCH 19pg, Hemoglobin electrophoresis show HbA of 88% and HbA2 of 4.2%.

Q.1) What is your diagnosis?  
Q.2) What is Mentzer Index?

ChatGPT  
Based on the clinical presentation and hemoglobin electrophoresis results:

1. The likely diagnosis is **"beta-thalassemia trait"**. Beta-thalassemia trait is a genetic condition characterized by reduced production of beta-globin chains, leading to microcytic hypochromic anemia. In individuals with beta-thalassemia trait, the hemoglobin electrophoresis typically shows elevated levels of hemoglobin A (HbA) with a normal or slightly increased level of hemoglobin A2 (HbA2). The presence of microcytic anemia (MCV of 55fL, MCH of 19pg) and elevated HbA (88%) along with normal HbA2 (4.2%) is consistent with the diagnosis of beta-thalassemia trait.
2. **"Mentzer Index"** is a simple calculation used to differentiate between iron deficiency anemia and beta-thalassemia trait based on the red blood cell indices. It is calculated by dividing the mean corpuscular volume (MCV) by the red blood cell count (RBC). The formula is as follows:
 
$$\text{Mentzer Index} = \text{MCV (fL)} / \text{RBC (million/mm}^3\text{)}$$

A Mentzer Index of less than 13 suggests iron deficiency anemia, while a Mentzer Index greater than 13 suggests beta-thalassemia trait. This index helps in distinguishing between these two types of microcytic anemia.

**Figure 3: Screenshot showing MCQ where answer selected was wrong but in explanation correct answer was explained.**

User  
A 42 year-old woman Paro presents with slowly progressive syndrome comprising of features like pain and tenderness in multiple joints, with joint stiffness on rising in the morning. Joint involvement is symmetric, with the proximal interphalangeal and metacarpophalangeal joints especially involved. The physician finds presence of tenderness in all the inflamed joints. Which of the following laboratory abnormalities is most likely associated in this patient?

(a) Antibodies to double-stranded DNA  
(b) IgM anti-IgG antibodies  
(c) HLA-B27 antigen  
(d) Urate crystals and neutrophils in synovial fluid

ChatGPT  
The clinical presentation described is characteristic of rheumatoid arthritis (RA), an autoimmune disease primarily affecting the joints. In RA, the most likely laboratory abnormality associated with the disease is:

(a) Antibodies to double-stranded DNA (anti-dsDNA antibodies).

However, it's worth noting that anti-dsDNA antibodies are more commonly associated with systemic lupus erythematosus (SLE), not RA. In RA, the presence of rheumatoid factor (RF) and anti-cyclic citrullinated peptide (anti-CCP) antibodies are more specific markers. So, the correct answer could be (b) IgM anti-IgG antibodies.

**Figure 4: Screenshot showing MCQ where the selected answer as well as the explanation was incorrect**

User

Which one of the following sequences correctly describes the usual temporal sequence of T-lymphocyte maturation within the cortex and medulla of the thymus?

- a. Double negative → double positive → single positive
- b. Double negative → single positive → double positive
- c. Double positive → double negative → single positive
- d. Double positive → single positive → double negative
- e. Single positive → double positive → double negative

ChatGPT

The correct sequence describing the usual temporal sequence of T-lymphocyte maturation within the cortex and medulla of the thymus is **\*\*b. Double negative → single positive → double positive\*\***.

In the thymus, T-lymphocytes undergo a maturation process involving several stages characterized by the expression of different surface markers. Initially, immature T-cells lack both CD4 and CD8 surface markers, making them "double negative" (CD4<sup>-</sup>CD8<sup>-</sup>). Then, they acquire either CD4 or CD8, becoming "single positive" (CD4<sup>+</sup>CD8<sup>-</sup> or CD4<sup>-</sup>CD8<sup>+</sup>). Finally, they express both CD4 and CD8, becoming "double positive" (CD4<sup>+</sup>CD8<sup>+</sup>).