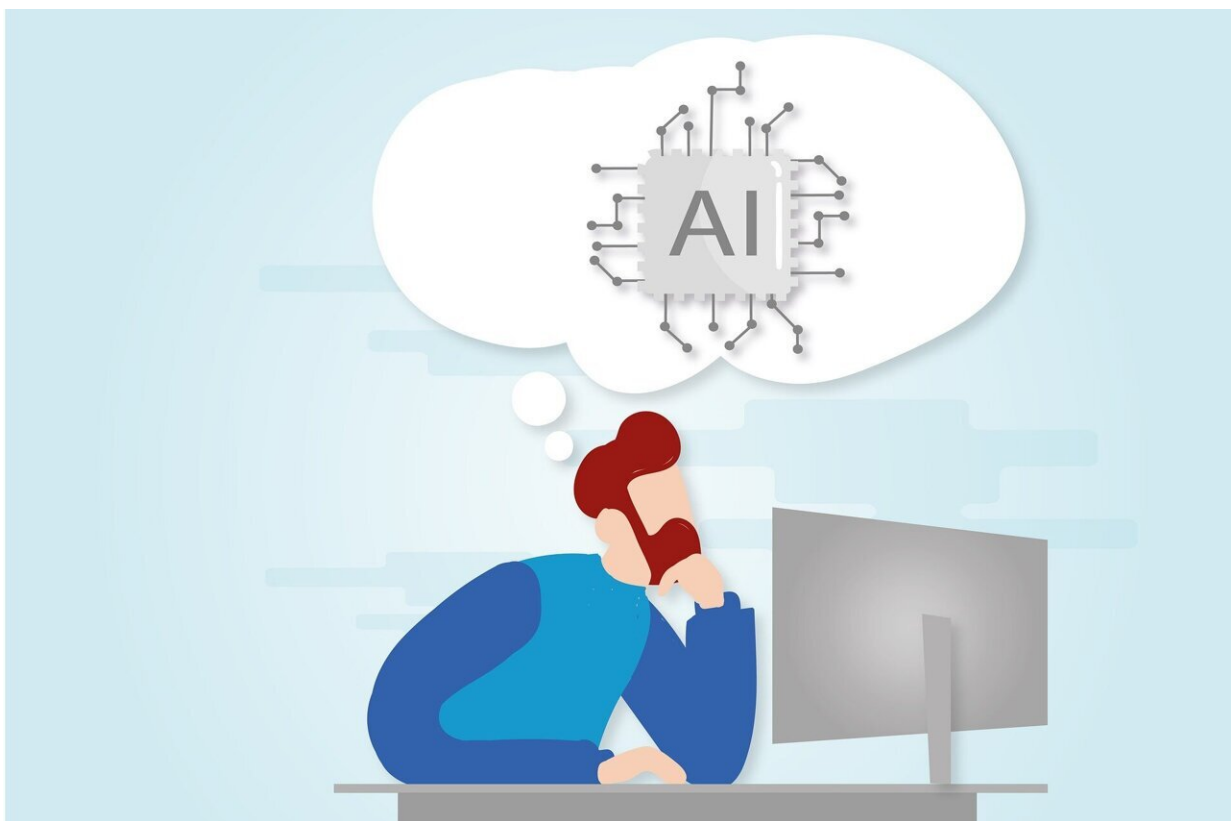# Study finds AI language model failed to produce appropriate questions, answers for medical school exam

December 20 2023

Credit: Pixabay/CC0 Public Domain

With concerns mounting that artificial intelligence (AI) could have a profound impact on traditional teaching in academic settings, many

question the role of ChatGPT, a sophisticated AI language model that can generate content that mimics human conversation.

ChatGPT offers the potential to assist or take over the student writing process with the capability of authoring everything from college admissions essays to term papers. But, can it also be used to aid the prodigious, sometimes daunting learning process in the medical school curriculum?

Researchers from Boston University Chobanian & Avedisian School of Medicine used ChatGPT to create multiple-choice questions, along with explanations of correct and incorrect choices, for a graduate and medical school immunology class that was taught by faculty in the school's department of pathology & laboratory medicine. They found the AI language model wrote acceptable questions but failed to produce appropriate answers.

The study is published in the journal *Academic Pathology*.

"Unfortunately, ChatGPT only generated correct questions and answers with explanations in 32% of the questions (19 out of 60 individual questions). In many instances, ChatGPT failed to provide an explanation for the incorrect answers. An additional 25% of the questions had answers that were either wrong or misleading," explained corresponding author Daniel Remick, MD, professor of pathology & laboratory medicine at the school

According to the researchers, students appreciate practice exams that can be used to study for their actual exams. These practice exams have even greater utility when explanations for answers are included since students will learn the rationale for the correct answer and have explanations for the incorrect answers.

Since ChatGPT generated questions with vague or confusing question stems and poor explanations of the answer choices, this study tool may not be entirely viable.

"These types of misleading questions may create further confusion about the topics, especially since the students have not gained expertise and they may not be able to find errors in the questions. "However, despite the issues we encountered, instructors may still find ChatGPT useful for creating practice exams with explanations—with the caveat that extensive editing may be required," added Remick.

More information: Alexander Ngo et al, ChatGPT 3.5 fails to write appropriate multiple choice practice exam questions, *Academic Pathology* (2023). DOI: 10.1016/j.acpath.2023.100099

Provided by Boston University School of Medicine