

عنوان مقاله:

Assessing ChatGPT's performance in national nuclear medicine specialty examination: An evaluative analysis

محل انتشار:

مجله پزشکی هسته ای ایران، دوره 32، شماره 1 (سال: 1403)

تعداد صفحات اصل مقاله: 6

نویسندگان:

Jakub Kufel - *Department of Biophysics, Faculty of Medical Sciences, Medical University of Silesia, Zabrze, Poland*

Michał Bielówka - *Professor Zbigniew Religa Student Scientific Association, Department of Biophysic, Faculty of Medical Sciences, Medical University of Silesia, Zabrze, Poland*

Marcin Rojek - *Professor Zbigniew Religa Student Scientific Association, Department of Biophysic, Faculty of Medical Sciences, Medical University of Silesia, Zabrze, Poland*

Adam Mitręga - *Professor Zbigniew Religa Student Scientific Association, Department of Biophysic, Faculty of Medical Sciences, Medical University of Silesia, Zabrze, Poland*

Łukasz Czogalik - *Professor Zbigniew Religa Student Scientific Association, Department of Biophysic, Faculty of Medical Sciences, Medical University of Silesia, Zabrze, Poland*

Dominika Kaczyńska - *Professor Zbigniew Religa Student Scientific Association, Department of Biophysic, Faculty of Medical Sciences, Medical University of Silesia, Zabrze, Poland*

Dominika Kondot - *Wielospecjalistyczny Szpital Powiatowy S.A. im. dr B. Hagera Pyskowicka ۴۷-۵۱, ۴۲-۶۱۲, Tarnowskie Góry, Poland*

Kacper Palkij - *Wielospecjalistyczny Szpital Powiatowy S.A. im. dr B. Hagera Pyskowicka ۴۷-۵۱, ۴۲-۶۱۲, Tarnowskie Góry, Poland*

Sylwia Mielcarska - *Department of Medical and Molecular Biology, Faculty of Medical Sciences, Medical University of Silesia, Zabrze, Poland*

خلاصه مقاله:

Introduction: The rapid development of artificial intelligence (AI) has sparked a desire to analyse its potential applications in medicine. The aim of this article is to present the effectiveness of the ChatGPT advanced language model in the context of the pass rate of the polish National Specialty Examination (PES) in nuclear medicine. It also aims to identify its strengths and limitations through an in-depth analysis of the issues raised in the exam questions.
Methods: The PES exam provided by the Centre for Medical Examinations in Łódź, consisting of ۱۲۰ questions, was used for the study. The questions were asked using the openai.com platform, through which free access to the GPT-۳.۵ model is available. All questions were classified according to Bloom's taxonomy to determine their complexity and difficulty, and according to two authors' subcategories. To assess the model's confidence in the

validity of the answers, each questions was asked five times in independent sessions. Results: ChatGPT achieved ۵۶%, which means it did not pass the exam. The pass rate is ۶۰%. Of the ۱۱۷ questions asked, ۶۶ were answered correctly. In the percentage of each type and subtype of questions answered correctly, there were no statistically significant differences. Conclusion: Further testing is needed using the questions provided by Centre for Medical Examinations from the nuclear medicine specialty exam to evaluate the utility of the ChatGPT model. This opens the .door for further research on upcoming improved versions of the ChatGPT

کلمات کلیدی:

Artificial intelligence, Computer science, Language model, Nuclear medicine exam

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/1897347>

