ISSN 2976-5498

ChatGPT versus Paediatricians: Who reigns supreme in paediatric exam success?*

Wilcox H[1], Davies T[2,3]

1. St. George's University Hospitals NHS Foundation Trust
2. Barts Health NHS Trust
3. Queen Mary University of London

Corresponding Author:
Full Name: Helena Wilcox
Email Address: Helena.wilcox1@nhs.net

London Paediatrics 2023: Volume 4

Accepted for publication October 2023

To subscribe to London Paediatrics:

http://journal.londonpaediatrics.co.uk

Article submissions and author instructions:

http://journal.londonpaediatrics.co.uk/index.php/1/about/submissions

ISSN 2976-5498

Abstract

Artificial intelligence (AI) is a rapidly progressing technology with increasing potential. This article aims to assess whether an AI programme is able to interpret questions from the MRCPCH, and whether it has the knowledge of a paediatrician. Questions from the MRCPCH were inputted into an AI programme and compared to answers given by the author, a paediatric trainee. Results showed an excellent understanding of paediatric medicine with potential for this to be utilised to aid clinicians in the future.

ISSN 2976-5498

Introduction

Artificial Intelligence (AI) is fast becoming integrated in all aspects of life, including medicine. It has great potential to facilitate clinicians as well as improve patient experience. As paediatricians, we have spent several years of our training revising for exams and have worked hard to understand the intricacies of paediatric medicine. With AI proving it can do most things, does it have the same knowledge as a paediatrician?

Method

This study's aims were to challenge an AI programme to answer questions from the MRCPCH examination to see if it has enough knowledge to pass the exam. This involved inputting 4 extended matching questions (EMQ) into ChatGPT from the MRCPCH Theory & Science examination specimen paper available via the RCPCH website.[1,2] Two versions of ChatGPT was used, version 3.5 and the latest 4.0.

Four EMQ's, each comprising of 3 questions, were submitted to ChatGPT. Instructions were given to only use one answer once. The test was also taken by the author, a paediatric trainee, to compare results.

The four subject domains were: science of practice; gastroenterology and hepatology; pharmacology and metabolic medicine.

The AI programmes was given instructions exactly as per the candidate paper, for example the following instructions were given for the first question (science of practice):

"You will be presented with three scenarios and a list of types of study design. Type in one answer only for each of the 3 scenarios below. Note: each answer may be used more than once."[1]

Results

The AI programme ChatGPT version 3.5 correctly answered 7 questions out of 12 (58%) whilst the author correctly answered 9 questions out of 12 (75%). In the first two questions ChatGPT 3.5 repeated the same answer, showing that it did not clearly understand the instructions as they were given in the paper. After this, ChatGPT 3.5 was prompted to only deliver one answer with the following:

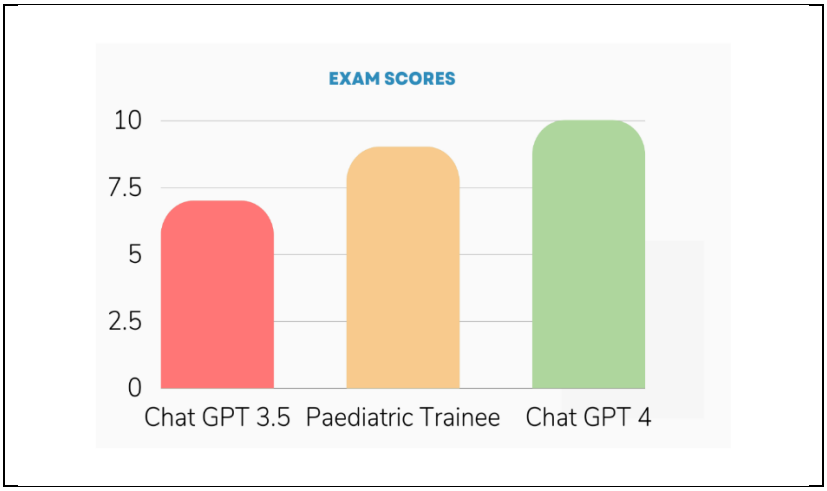

Figure 1. *Exam Scores determined by trainee and Chat GPT.*

Author: remember you can only choose one answer once

ChatGPT: Understood, I will provide one answer for each of the three scenarios without repeating any answers.

ISSN 2976-5498

Of all the domains, chat GPT 3.5 scored the least points in the metabolic medicine questions, receiving 1/3 whilst the author received 3/3 points.

When using the latest version of ChatGPT 4.0, the AI programme got 10 questions correct giving it an overall score of 83%. It did not repeat the same answer twice, showing that it had the ability to understand the initial instructions.

Discussion

Although chatGPT version 3.5 seemed to have reasonable knowledge, it did not initially understand the basic task of not repeating the answers for each of the 3 clinical scenarios until it was prompted to do so. Results were low for certain subjects, showing that more of an understanding of paediatric medicine is required for the MRCPCH. However, the latest version ChatGPT 4.0 was able to beat the paediatric trainee's mark as well as understand the exam instructions without further prompting. The format of the questions required the candidate to initially recognise the condition, and to then know which investigation was necessary to correctly answer the questions. This may have proved to be difficult for ChatGPT 3.5 however ChatGPT 4.0 was able to interpret this and provide a correct answer.

Limitations of the study were that only 12 questions in total were used and there are many other areas of paediatric medicine that could have been covered.

It is worth noting that both versions of the AI programme had not been given any further information than what it knew already. As such, there is potential to improve the score with further training. This could involve giving it information directly from MRCPCH revision textbooks. Both AI programmes took a matter of seconds to reach their answer for each individual question whereas the paediatric trainee took a total of 30 minutes to answer the paper in its entirety. This shows the speed at which this technology can work. This study shows the advancements in AI technology, even between two versions of the same programme. It demonstrates that AI has the ability to pass an area of the MRCPCH at the level of a paediatrician. With the technology rapidly improving, it is worth exploring how we can use it to our advantage and how it may assist clinicians.

*The title of this article was generated by ChatGPT.

References

1. Royal College of Paediatrics & Child Health. Theory Exam Sample Papers. Taken from https://www.rcpch.ac.uk/resources/theory-exam-sample-papers
2. Royal College of Paediatrics & Child Health. Theory & Science Specimen Paper – answers and reasoning. Taken from https://www.rcpch.ac.uk/sites/default/files/2019-07/TAS%20Specimen%20paper%20-%20answer%20key%20and%20reasoning.pdf