Comparing AI Performance with Ophthalmology Resident Doctors: A Standardised Assessment of ChatGPT-4o

Zaid Alsafi[1], Mohaimen Al-Zubaidy[2], Somain Verma[3]

1. Western Eye Hospital, Imperial College Healthcare NHS Trust, UK 2. Royal Victoria Infirmary, The Newcastle Upon Tyne Hospitals NHS Foundation Trust, UK 3. Leicester Royal Infirmary, University Hospitals of Leicester NHS Trust

Background:

- Artificial intelligence (AI) is increasingly integrated into medical education, providing personalised learning opportunities and efficient knowledge assessment.
- ChatGPT-4o, released in May 2024, represents the latest and most advanced iteration of large language models.

Aim:

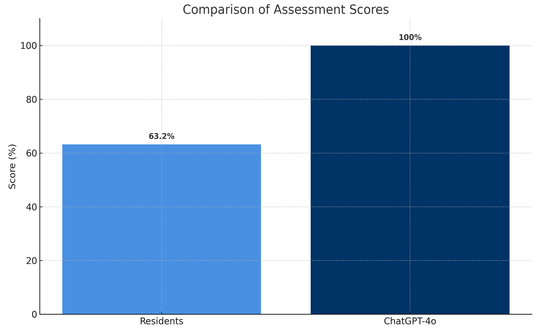
- This study aimed to compare the performance of ChatGPT-4o with that of ophthalmology trainees who had recently passed the Part 1 Fellowship Examination of the Royal College of Ophthalmologists (FRCOphth), using a standardised set of multiple-choice questions (MCQs).

Method:

- A cohort of 15 trainees, ranging from Foundation Year 1 (FY1) to Ophthalmic Specialty Trainee Year 2 (ST2), each of whom had passed the Part 1 FRCOphth examination within the preceding years, were recruited.
- Participants completed a set of 50 MCQs sourced from the Royal College of Ophthalmologists' sample question bank.
- Their scores were recorded and subjected to statistical analysis to determine the mean and range.
- ChatGPT-4o was subsequently tasked with answering the identical MCQ set, and its responses were evaluated against the official answer key.
- The performance of ChatGPT-4o was then compared directly to that of the resident cohort.

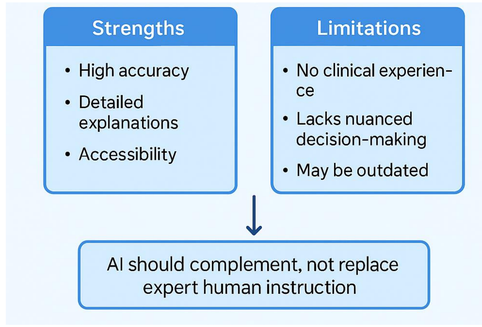
Results:

- The residents achieved a mean score of 31.6 out of 50 (63.2%), with individual scores ranging from 23 to 41 (46% to 82%).
- By contrast, ChatGPT-4o attained a perfect score of 50 out of 50 (100%).
- For additional context, the average pass mark for the Part 1 FRCOphth examination over the preceding year was approximately 62.4%.



Discussion:

- ChatGPT-4o, released in May 2024, achieved perfect performance in a standardized ophthalmology MCQ exam, outperforming trainees.
- Its ability to deliver accurate, well-explained answers suggests promise as an adjunctive learning tool in ophthalmic education.
- However, its use must be tempered with caution:



- Ongoing research is needed to evaluate its impact on knowledge retention, clinical skills, and patient care.