

Əlihüseynli E.H.

PEDIATRİK OFTALMOLOGİYA VƏ ÇƏPGÖZLÜK ÜZRƏ ÇOX SEÇİMLİ SUALLARIN CAVABLANDIRILMASINDA BÖYÜK DİL MODELLƏRİNİN VƏ AZƏRBAYCAN DİLLİ TERMİNOLOGİYAYA UYGUNLAŞDIRILMIŞ CHATGPT ƏSASLI “OFTALMOASSİST”İN MÜQAYİSƏLİ QİYMƏTLƏNDİRİLMƏSİ

Müasir Oftalmologiyanın İnkişafına Yardım Birliyi

Sumqayıt, Azərbaycan

E-mail: eminalihuseynli@gmail.com

<https://www.doi.org/>

Giriş

Böyük dil modellərinin (BDM) rezidentlərin tibbi təhsilində istifadəsi son illərdə geniş maraq doğurur. Pediatrik oftalmologiya və çəpgözlük sahəsində bu modellərin effektivliyi yalnız klinik biliklərlə deyil, eyni zamanda terminoloji dəqiqlik və dil uyğunluğu ilə müəyyən olunur [1 -5].

Məqsəd – pediatrik oftalmologiya və çəpgözlük üzrə çox seçimli sualları (ÇSS) cavablandırmada aparıcı BDM-in və Azərbaycan dilli oftalmoloji terminologiyaya uyğunlaşdırılmış ChatGPT əsaslı “Oftalmoassist” modelinin rezident təhsilində potensial rolunu müqayisəli şəkildə qiymətləndirmək.

Material və metodlar

Tədqiqatda oftalmologiya üzrə beynəlxalq sertifikatlaşdırma imtahanlarından götürülmüş, pediatrik oftalmologiya və çəpgözlük sahəsini əhatə edən ÇSS istifadə olunmuşdur. Suallar kliniki yanaşma, terminoloji uyğunluq və vizual materialdan istifadə meyarlarına görə kateqoriyalaşdırılmışdır. Grok 4.1, Gemini-3, DeepSeek-4.1, standart ChatGPT və Azərbaycan dilli terminologiyaya uyğunlaşdırılmış ChatGPT əsaslı “Oftalmoassist” modelləri sınaqdan keçirilmişdir. Dəqiqlik göstəriciləri faizlə ifadə olunmuş və müqayisəli təhlil aparılmışdır.

Nəticələr

Ümumilikdə bütün BDM pediatrik oftalmologiya və çəpgözlük üzrə ÇSS-lərin cavablandırılmasında yüksək nəticə göstərmişdir. Modellər arasında DeepSeek klinik məntiq tələb edən suallarda orta hesabla 85-88%, Gemini isə 80-83% dəqiqlik nümayiş etdirmişdir. Grok xüsusilə nəzəri biliklərə əsaslanan suallarda 82-85%, standart ChatGPT modeli isə 88-89% aralığında nəticələr göstərmişdir. Azərbaycan dilli terminologiyaya uyğunlaşdırılmış “Oftalmoassist” isə ümumi dəqiqlik baxımından bütün modelləri üstələmiş, 90%-dən artıq cavab uyğunluğu əldə etmişdir. Pediatrik oftalmologiya və çəpgözlük terminologiyası və klinik ssenariyə əsaslanan suallarda “Oftalmoassist” 92-94% dəqiqliklə stabil və kontekstə uyğun nəticələr göstərmişdir. Hesablama tələb edən və etməyən

suallar üzrə “Oftalmoassist”in nəticələri arasında əhəmiyyətli fərq müşahidə olunmamışdır (93,1% və 92,4%), bu isə modelin balanslı bilik strukturuna malik olduğunu göstərir.

Yekun

Əldə edilən nəticələr göstərir ki, BDM-lər pediatrik oftalmologiya və çəpgözlük üzrə rezident təhsilində yüksək potensiala malikdir. Bununla yanaşı, Azərbaycan dilli oftalmoloji terminologiyaya uyğunlaşdırılmış ChatGPT əsaslı “Oftalmoassist” terminoloji dəqiqlik, klinik uyğunluq və ümumi cavabvermə dəqiqliyi baxımından digər modellərlə müqayisədə daha üstün nəticə nümayiş etdirmişdir. Bu üstünlük “Oftalmoassist”in yerli tibbi təhsil mühitində imtahana hazırlıq və tədris dəstəyi üçün xüsusi əhəmiyyət daşıdığını göstərir. Gələcək tədqiqatlarda daha geniş sual bazaları və çoxmərkəzli yanaşmalarla bu nəticələrin təsdiqlənməsi məqsədəuyğun hesab olunur.

Açar sözlər: pediatrik oftalmologiya, çəpgözlük, böyük dil modelləri, “Oftalmoassist”, ChatGPT, çox seçimli suallar

Alihuseynli E.H.

COMPARATIVE EVALUATION OF LARGE LANGUAGE MODELS AND THE CHATGPT-BASED “OFTALMOASSIST” ADAPTED TO AZERBAIJANI OPHTHALMIC TERMINOLOGY IN ANSWERING MULTIPLE-CHOICE QUESTIONS IN PAEDIATRIC OPHTHALMOLOGY AND STRABISMUS

Assistance for Development of Modern Ophthalmology Union

Sumgayit, Azerbaijan

E-mail: eminalihuseynli@gmail.com

<https://www.doi.org/>

Introduction

The use of Large Language Models (LLMs) in medical education for residents has attracted increasing interest in recent years. In paediatric ophthalmology and strabismus, the effectiveness of these models is determined not only by clinical knowledge but also by terminological precision and language compatibility [1 -5].

Purpose – to comparatively evaluate the performance of leading LLMs and the ChatGPT-based model “Oftalmoassist” adapted to Azerbaijani ophthalmological terminology, in answering multiple-choice questions (MCQs) in paediatric ophthalmology and strabismus, and to assess their potential role in resident education.

Material and Methods

The study utilized MCQs covering paediatric ophthalmology and strabismus obtained from international ophthalmology certification examinations. Questions were categorized according to clinical reasoning requirements, terminological accuracy, and the use of visual materials. The models tested included Grok 4.1, Gemini-3, DeepSeek-4.1, standard ChatGPT, and the ChatGPT-based “Oftalmoassist,” adapted to Azerbaijani ophthalmic terminology. Accuracy rates were expressed as percentages and comparatively analysed.

Results

Overall, all LLMs demonstrated strong performance in answering MCQs in paediatric ophthalmology and strabismus. Among the models, DeepSeek achieved average accuracy rates of 85–88% in questions requiring clinical reasoning, while Gemini demonstrated 80-83% accuracy. Grok performed particularly well in theory-based questions, achieving 82-85% accuracy, whereas the standard ChatGPT model showed results in the range of 88-89%. The Azerbaijani terminology-adapted model “Oftalmoassist” outperformed all other models in overall accuracy, exceeding 90% answer concordance. In terminology-sensitive and clinical scenario-based paediatric ophthalmology and strabismus questions, “Oftalmoassist” demonstrated stable and contextually appropriate performance with accuracy rates of 92–94%. No significant difference was observed between questions requiring calculations and those not requiring calculations (93.1% vs. 92.4%), indicating a balanced knowledge structure within the model.

Conclusion

The findings indicate that LLMs have substantial potential in resident education in pediatric ophthalmology and strabismus. However, the ChatGPT-based “Oftalmoassist” model, adapted to Azerbaijani ophthalmic terminology, demonstrated superior performance compared to other models in terms of terminological precision, clinical relevance, and overall response accuracy. This advantage highlights its particular value for exam preparation and educational support within the local medical training environment. Future studies involving larger question banks and multicenter approaches are recommended to further validate these findings.

Keywords: *pediatric ophthalmology, strabismus, Large Language Models, “Oftalmoassist”, ChatGPT, multiple-choice questions*

ƏDƏBİYYAT | REFERENCES

1. Sawamura, S. *An Evaluation of the Performance of OpenAI-o1 and GPT-4o in the Japanese National Examination for Physical Therapists* / S.Sawamura, K.Kohiyama, T.Takenaka [et al.] // *Cureus*, – 2025. Jan; 6. 17(1), – p. 76989. <https://doi.org/10.7759/cureus.76989>
2. Patil, A. *Advancing Reasoning in Large Language Models: Promising Methods and Approaches* / A.Patil, A.Jadon // *arXiv*, – 2025. 2502, – p. 03671. <https://doi.org/10.48550/arXiv.2502.03671>
3. Goh, E. *Large Language Model Influence on Diagnostic Reasoning: A Randomized Clinical Trial* / E.Goh, R.Gallo, J.Hom [et al.] // *JAMA Netw Open*, – 2024. Oct; 1. 7(10), – p. 2440969. <https://doi.org/10.1001/jamanetworkopen.2024.40969>
4. Aung, Y.Y.M. *The promise of artificial intelligence: a review of the opportunities and challenges of artificial intelligence in healthcare* / Y.Y.M.Aung, D.C.S.Wong, D.S.W.Ting // *Br. Med. Bull.*, – 2021. 139, – p. 4-15. <https://doi.org/10.1093/bmb/ldab016>
5. Mistry, N.P. *Large Language Models as Tools to Generate Radiology Board-Style Multiple-Choice Questions* / N.P.Mistry, H.Saeed, S.Rafique [et al.] // *Acad. Radiol.*, – 2024. Sep; 31(9), – p. 3872-3878. <https://doi.org/10.1016/j.acra.2024.06.046>