

Editöre Mektup/Letter to the Editor

European Diploma in Breast Imaging Sample Multiple Choice Questions: Evaluating the Capabilities of the Large Language Models

Avrupa Meme Görüntüleme Diploması Çoktan Seçmeli Örnek Soruları: Büyük Dil Modellerinin Yeteneklerinin Değerlendirilmesi

Muhammed Said BEŞLER¹

Dear Editor,

In pursuit of a fresh outlook, my objective was to gauge the abilities of a multimodal large language model (LLM) against sample questions from the European Diploma in Breast Imaging (EDBI) test, an initiative by the European Society of Breast Imaging.

Large language models are pushing the potential in radiology, from interpreting text and medical images to generating reports (Bhayana, 2024). Generative Pre-trained Transformer 4 (GPT-4) has notably passed a national mammography board exam with clarity (Almeida et al, 2024). As the latest version among multimodal LLM types, GPT-4 is capable of answering questions requiring both lower-order and higher-order thinking. Three written sample questions, where multiple choices could be correct, were evaluated. It was noted that there was no negative marking for incorrect answers (<https://www.eusobi.org/european-diploma-in-breast-imaging-edbi/>). The scoring system was adapted from the European Diploma in Radiology scoring guidelines (<https://www.myebr.org/edir-scoring-faqs>). Data was obtained from Google Gemini, GPT-3.5, and GPT-4 in March 2024. When assigning a value of 1 point to each question, GPT-4 reached an accuracy of 78%, GPT-3.5 achieved 50%, and Google Gemini scored 22.2%. This notable success in the sample questions from the EDBI particularly emphasizes GPT-4's potential in aiding clinical decision-making. Future studies may assess its performance in questions requiring medical image analysis, such as mammography, breast ultrasound, or breast magnetic resonance imaging.

¹Uzman Dr., Kahramanmaraş Necip Fazıl Şehir Hastanesi, ORCID: 0000-0001-8316-7129, msbesler@gmail.com
M.D., Kahramanmaraş Necip Fazıl City Hospital

References

- Almeida, L.C., Farina, E. M. J. M., Kuriki, P. E. A., Abdala, N., & Kitamura, F. C (2024). Performance of ChatGPT on the Brazilian Radiology and Diagnostic Imaging and Mammography Board Examinations. *Radiol Artif Intell*, 6(1), e230103. <http://dx.doi.org/10.1148/ryai.230103>
- Bhayana, R. (2024). Chatbots and Large Language Models in Radiology: A Practical Primer for Clinical and Research Applications. *Radiology*, 310(1), e232756. <http://dx.doi.org/10.1148/radiol.232756>