

## Comparison of the Performance of ChatGPT-3.5, ChatGPT-4o and Gemini 2.5 in the Management of Rheumatoid Arthritis: Analysis of EULAR 2022 Recommendations

Ozlem Kuculmez

Ahmet Usen

### Abstract

**Background:** Large language models such as ChatGPT and Gemini have become widely popular due to their fast responses. The aim of this research is to assess and contrast the answers provided by different large language models to open-ended questions based on the 2022 EULAR guideline.

**Materials and Methods:** This was a comparative, cross-sectional study. Five overarching principles and eleven recommendations from the 2022 EULAR guidelines were rephrased into open-ended clinical questions. These questions were presented to ChatGPT-3.5, GPT-4o, and Gemini 2.5 Flash. The responses were assessed using a 7-point Likert scale for usability and reliability, readability scores (Flesch-Kincaid Reading Ease (FKRE) and Flesch-Kincaid Grade Level (FKGL)), and similarity metrics (Universal Sentence Encoder (USE) and ROUGE-L). Different models were compared statistically and  $p < 0.05$  was revealed as statistically significant.

**Results:** ChatGPT-3.5 and ChatGPT-4.0 had significantly higher FKGL and FKRE scores than Google Gemini 2.5 ( $p < 0.05$ ). Reliability did not differ between models ( $p > 0.05$ ), but ChatGPT-4.0 had a significantly higher usefulness score than Gemini 2.5 ( $p < 0.05$ ). ChatGPT-4o and Gemini 2.5 showed significantly higher semantic similarity than ChatGPT-3.5 ( $p < 0.05$ ). ROUGE-L F1 scores showed no significant differences ( $p > 0.05$ ). FKRE and FKGL were strongly negatively correlated, while reliability and usefulness were strongly positively correlated ( $p < 0.05$ ).

**Conclusions:** It was determined that ChatGPT-3.5 and ChatGPT-4o programs gave more readable, and useful answers to open-ended questions derived from the EULAR 2022 guideline, although there was no difference in reliability. Further studies are needed.

### Keywords

Arthritis, artificial intelligence, guideline, rheumatic disease, rheumatoid arthritis.

