The tutelage of ChatGPT: Assessing AI's capability to simplify radiology physics, a bilingual approach.

Shardul Tyagi [1], Namrita Sachdev [1]
[1] ABVIMS and Dr. RML Hospital, New Delhi, India.

1.Introduction

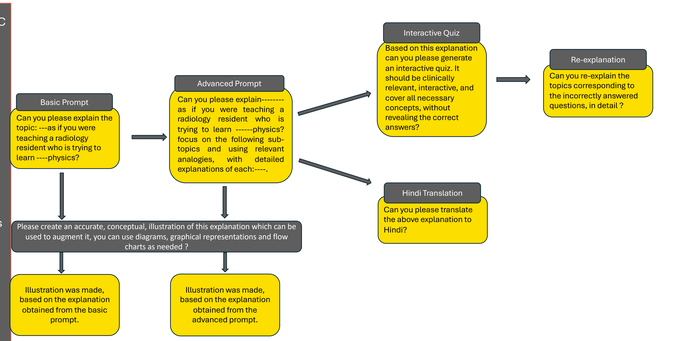
Artificial intelligence (AI) is the simulation of human cognitive capacities using computer programming, allowing computer programs to emulate human thought and behaviour. ChatGPT based on the generative pre-trained transformer architecture is designed to generate human like responses to text based input. This has led to a massive interest in its varied applications, education highlighting a major use case for this technology. The field of radiology is at the frontier of cutting medical innovation, and AI has the potential to be effectively inculcated at almost every step of the process, from radiology training to patient care.

2. Aims & Objectives

- Our study aims to evaluate the effectiveness of AI tools, specifically GPT 4.0 in radiology education.
- As other studies have highlighted the many general use cases of ChatGPT, our study aims to evaluate its performance pertaining to a specific task i.e., teaching radiology physics.
- Our study is also a step towards answering the bigger question about the inculcation of AI based tools in radiology training.

3. Materials and Methods

- Study design- Cross-sectional observational; conducted using GPT 4.0, for 55 topics across NMC radiology physics curriculum.
- General Explanations for each topic were obtained by GPT-4 first responding to a standardised basic prompt followed by an advanced prompt with a focus on specific sub-topics and use of analogies.
- Following this Hindi translations, illustrations, interactive quizzeas and re-explanations wee generated as shown in the flow chart.
- Two board certified radiologists assessed these using individually adapted 5-point Likert scales, Translations were assessed on Accuracy (Acu), Consistency (Con), Fidelity (F) and Comprehensibility (Comp) while the rest were assessed on Accuracy (Acu), Comprehensiveness (Com), Didactic Value (DV) and Engagement (En).
- Data were analysed using SPSS V.29, inferential statistics using unpaired T-test were performed to do comparative analysis of mean Likert scale scores. Explanations and illustrations obtained from using basic and advanced prompts were compared. Hindi translations, quizzes and re-explanations were individually assessed.



4. Results

- Advanced prompts resulted in explanations that were significantly better ($p<0.05$) in Acu, Com, DV and En as compared to the basic prompts. Of note is the incorporation of analogies in the prompt, which lead to significantly better explanations.
- The illustrations obtained using both basic and advanced prompts though being engaging En- 3.45±0.50, in terms of the design, performed poorly on other assessment parameters, with mean scores of- Acu- 1.82±0.6; Com- 2.24±0.57; DV- 1.56±0.50. They were inaccurate, being riddled with spelling errors, at times using incomprehensible language.
- Hindi translations performed moderately on all parameters with mean scores shown in fig. 3. The translations though being accurate where in parts, inconsistent, lacking fidelity to the English explanation.
- The interactive quizzes scored highly on accuracy and moderately on the remaining parameters (fig. 4). Their limitation being the scope of the questions, where at times even the analogies used in the explanations were presented as direct questions.
- Re-explanations scored highly on all parameters with the increase likely pertaining to the specific focus on limited topics.

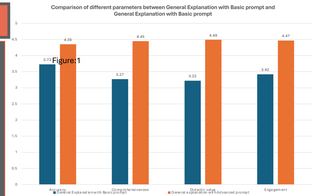


Figure:1

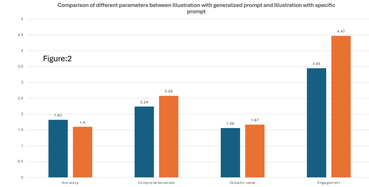


Figure:2



Figure:3 Hindi translation characteristics



Figure:4 Interactive Quiz

5. Conclusion

- Our study shows that GPT 4 generated explanations, Interactive quizzes and re-explanations were accurate, comprehensive, engaging and educational. Illustrations though engaging were highly inaccurate with no supplementary value.
- Hindi translations were accurate but inconsistent highlighting the AI's linguistic limitations.
- In conclusion GPT 4 can be an effective learning tool for basic radiology physics.
- Further research using different applications are needed to pave way for the incorporation of AI into radiology training. This can potentially have a dual impact, improving upon the existing training and decreased friction in adoption of the technology.

References

1.Sodhi KS, Tao TY, Seymore N. ChatGPT: Chasing the Storm in Radiology Training and Education. Indian J Radiol Imaging. 2023 Oct 5;33(4):431-435.

2.Grewal H, Dhillon G, Monga V, Sharma P, Buddhavarapu VS, Sidhu G, Kashyap R. Radiology Gets Chatty: The ChatGPT Saga Unfolds. Cureus. 2023 Jun 8;15(6):e40135.

3.Scheschenja M, Viniol S, Bastian MB, Wessendorf J, König AM, Mahnken AH. Feasibility of GPT-3 and GPT-4 for in-Depth Patient Education Prior to Interventional Radiological Procedures: A Comparative Analysis. Cardiovasc Intervent Radiol. 2024 Feb;47(2):245-250.

4. Meşe İ, Taşlıçay CA, Kuzan BN, Kuzan TY, Sivrioğlu AK. Educating the next generation of radiologists: a comparative report of ChatGPT and e-learning resources. Diagn Interv Radiol . 2024 May;30(3):163-174.

5.Perera Molligoda Arachchige AS. Comment: ChatGPT: Chasing the Storm in Radiology Training and Education. Indian J Radiol Imaging. 2024 Feb 6;34(3):569-570.