Title: **Evaluating ChatGPT-4o in Echocardiographic Interpretation: Accuracy in Identifying Imaging Modalities and Cardiac Structures**

Authors: Rudy R Unni, Chi-Ming Chow
Institution: St. Michael's Hospital

Background:
Recent advancements in artificial intelligence (AI), particularly large language models (LLMs) with integrated vision capabilities, have raised interest in their potential to assist in echocardiographic interpretation. OpenAI's ChatGPT-4o incorporates multimodal functionality, enabling image recognition. However, its ability to interpret echocardiographic images remains untested. This study evaluated the accuracy of ChatGPT-4o in identifying echocardiographic imaging modalities and cardiac structures.

Methods:
This prospective observational study was conducted at the Echocardiography Lab of St. Michael's Hospital, Toronto. We selected normal echocardiograms from 10 anonymized patients and identified images of M-Mode at the level of the left ventricle, mitral valve (MV), and aortic valve (AV), continuous wave Doppler across the AV, pulsed-wave Doppler of the MV inflow, and tissue doppler imaging (TDI) of the medial mitral annulus. Each image was uploaded to Chat GPT 4.o model, along with a prompt asking for the type of ultrasound imaging and structure, and whether the structures demonstrated normal or abnormal function. Responses were graded as correct or incorrect by senior echocardiography fellows/staff. A custom GPT was then developed by uploading the ASE 2019 Comprehensive TTE Guidelines, ASE 2016 Diastolic Function Guidelines, and ASE 2015 Chamber Quantification Guidelines for background training. The same test images from Phase 1 were re-evaluated post-training.

Results:
ChatGPT was accurate (60/60, 100%) at identifying the type of modality of US imaging with lower performance (52/60, 86%) at identifying the structures of interest. Performance was weakest at identifying M-mode through the mitral valve (5/10, 50%) versus through aortic valve (7/10, 70%). When oriented to the structure of choice and modality, GPT was able to identify normal studies as well as potentially abnormal function (59/60, 98%). After training of a custom GPT using ASE guidelines, the GPT performance improved in identifying the structure of choice (58/60, 97%). Different valve pathologies were also used to test the custom trained GPT with varying performance.

Conclusion:
ChatGPT-4o demonstrated good baseline accuracy in recognizing echocardiographic imaging modalities and cardiac structures, with improvement following structured training. These findings suggest AI could serve as an assistive tool for echocardiographic education and preliminary image interpretation. Further research is warranted to assess its integration into clinical practice.