Skeletal Radiology (2023) 52:1755–1758
https://doi.org/10.1007/s00256-023-04340-5

SCIENTIFIC ARTICLE

A comparison of ChatGPT-generated articles with human-written articles

Sisith Ariyaratne[1] · Karthikeyan. P. Iyengar[2] · Neha Nischal[3] · Naparla Chitti Babu[4] · Rajesh Botchu[1]

Received: 1 March 2023 / Revised: 6 April 2023 / Accepted: 9 April 2023 / Published online: 14 April 2023

Abstract

Objective ChatGPT (Generative Pre-trained Transformer) is an artificial intelligence language tool developed by OpenAI that utilises machine learning algorithms to generate text that closely mimics human language. It has recently taken the internet by storm. There have been several concerns regarding the accuracy of documents it generates. This study compares the accuracy and quality of several ChatGPT-generated academic articles with those written by human authors.

Material and methods We performed a study to assess the accuracy of ChatGPT-generated radiology articles by comparing them with the published or written, and under review articles. These were independently analysed by two fellowship-trained musculoskeletal radiologists and graded from 1 to 5 (1 being bad and inaccurate to 5 being excellent and accurate).

Results In total, 4 of the 5 articles written by ChatGPT were significantly inaccurate with fictitious references. One of the papers was well written, with a good introduction and discussion; however, all references were fictitious.

Conclusion ChatGPT is able to generate coherent research articles, which on initial review may closely resemble authentic articles published by academic researchers. However, all of the articles we assessed were factually inaccurate and had fictitious references. It is worth noting, however, that the articles generated may appear authentic to an untrained reader.

Keywords ChatGPT · Articles · Accuracy · Research

Introduction

ChatGPT is an artificial intelligence (AI) language tool developed by OpenAI that utilises machine learning algorithms to generate text that closely mimics human language [1]. ChatGPT has been trained on large amounts of data to improve its ability to understand and generate natural-language text and can comprehend and respond to natural-language inputs. A tool of this calibre has the ability to influence a diverse array of sectors. In the field of research, ChatGPT has demonstrated the ability to generate scientific papers that are similar to authentic papers written by academic researchers, which has raised many questions about its potential role in the future of academic research. Several concerns, however, have been raised regarding the accuracy of research texts generated by ChatGPT [2, 3]. This study aims to assess the accuracy of radiology research articles generated by ChatGPT by performing a comparison with authentic, published articles or those under review.

Materials and methods

Five randomly selected articles published prior to 2021 or written and under review by senior authors were selected. ChatGPT was asked to write the articles with references. ChatGPT version 3.0 was used. These were then independently compared by two fellowship-trained musculoskeletal radiologists with the published articles. The references generated were cross-referenced with scientific databases (PubMed, Google, and Ovid Medline) to assess for authenticity.

✉ Rajesh Botchu
drbrajesh@yahoo.com

[1] Department of Musculoskeletal Radiology, The Royal Orthopedic Hospital, Bristol Road South, Northfield, Birmingham, UK

[2] Department of Orthopedics, Southport and Ormskirk Hospital, Southport, UK

[3] Department of Radiology, Holy Family Hospital, New Delhi, India

[4] Department of Radiology, Srinivas Institute of Medical Sciences & Research Centre, Mukka, Mangalore, India