Qeios

Peer Review

Review of: "The Influence of an Artificial Intelligence Large Language Model (ChatGPT) on Orthopaedic Scientific Publishing: A Bibliometric Analysis"

Abuzer Güngor[1]

1. Istinye University, Istanbul, Turkey

The study looks at changes in orthopaedic publications before and after the release of ChatGPT. The dataset is large, and the topic is important, but the paper often makes stronger claims than the data can support. Many parts of the discussion go far beyond the actual results, and some conclusions rely on assumptions rather than evidence.

1. The biggest issue is the time comparison. The authors compare 2021–2022 with 2022–2025, but they do not show whether similar growth already existed in earlier years. Orthopaedic publishing has been increasing for a long time, independent of AI, and this must be checked. If publication numbers were already rising between 2017–2021 at a similar rate, then the observed increase cannot be attributed to ChatGPT. Right now, the study risks confusing a normal long-term trend with an AI effect.
2. The part about the number of authors per paper is not convincing. Tools like ChatGPT would be expected to *reduce* the need for many co-authors, not increase it. Growth in author numbers is more likely due to increasing collaborations, multi-centre work, or general changes in scientific culture. Claiming that ChatGPT influences author count is speculative and should be removed or rewritten more carefully.
3. The discussion spends too much time on general AI topics that are not directly related to the study's findings. A better discussion would focus on the actual data instead of repeating broad arguments about ethics, AI in medicine, or future scenarios. The paper should clearly state what the results show and what they do *not* show.

doi.org/10.32388/30M1FR

4. Another suggestion is to analyse articles that include an “AI declaration” or tag over the years. This could provide more direct evidence of AI adoption. Right now, the study measures indirect signals (sentence length, TTR, sentiment) that can be influenced by many factors beyond ChatGPT.
5. Some language changes are statistically significant but extremely small. For example, a 0.3-word difference in sentence length or a TTR change of 0.519 → 0.523 cannot be interpreted as a meaningful shift in writing style. The paper should avoid overstating these effects.

The study design itself is interesting and worth publishing. These changes would make the paper more honest, clearer, and actually useful for the field.

Reviewer.

Declarations

Potential competing interests: No potential competing interests to declare.