LETTER TO THE EDITOR
(CC BY-SA)

UDC: 311:61]:004.8
DOI: https://doi.org/10.2298/VSP261007065R

Impact of artificial intelligence on statistical analyses in biomedical research

Uticaj veštačke inteligencije na statističke analize u biomedicinskim istraživanjima

To the Editor

Artificial intelligence (AI) is increasingly being integrated into various scientific fields, particularly medicine and biomedical research [1]. One of the reasons is the precision of neural networks or machine learning that characterizes AI [2]. AI is rapidly transforming modern medicine, precisely with the aforementioned highly precise algorithms [3].

Among the early applications of AI in medicine was the interpretation of radiological images, while its use has since expanded to other fields, including surgery and robotic-assisted surgical procedures [4–6]. As AI advances, the limitations of AI-based decisions are becoming increasingly apparent, leading to ethical dilemmas and conflicting opinions that can be a significant problem in medical practice. The same is true of statistical analysis and interpretation of results. Any misinterpretation of results can have far-reaching consequences.

Statistical data analysis is not outside the scope of AI. AI can be used both for data analysis and statistical results, as well as for analyzing already done statistics and finding statistical errors [7]. Large language models (LLMs), such as Chat Generative Pre-Trained Transformer (ChatGPT) and Gemini, have shown impressive capabilities in text generation and in generating R code [8]. R codes are computer scripts written in the R programming language for statistical analysis. One study demonstrated that, although LLMs are capable of generating correct and interpretable R code in many cases, careful human oversight remains essential when integrating AI into statistical analyses to minimize the risk of misuse [3]. Another study showed that generative AI tools, such as ChatGPT, can facilitate data analysis; however, human oversight and adequate statistical knowledge remain essential [9].

The appropriate selection of statistical tests is crucial for ensuring the validity and reliability of research findings [10]. Errors, such as using the wrong test or misinterpreting data, can easily lead to incorrect conclusions in a study [11]. The most common errors in statistical analyses that lead to inaccurate results and misleading conclusions are errors in statistical design, data presentation, data outlier treatment, the use of parametric and nonparametric tests, the disclosure of post-hoc analyses, and the definition of the hypotheses [7]. Therefore, by identifying statistical errors, AI systems increase the reliability of research. AI helps non-experts analyze data, but it can be unpredictable for experts dealing with complex data analysis. Recent studies show that AI is increasingly helpful in research, assisting in spotting errors in methodology, citations, and statistical analyses. AI can work independently, but it works better if it is managed by someone with statistical knowledge [7]. AI also brings risks like misreading results, ethical issues, and privacy concerns, so researchers must make final decisions.

About two years ago, a paper was published in *Nature* discussing the use of AI in reviewing paper text and detecting errors in methodology, calculations, and inconsistencies in the citation of papers and statistical analysis [12]. Similar concerns arise when AI is used for manuscript drafting and data analysis, where errors generated by AI may require careful human verification [13]. There are also papers that focus exclusively on the analysis of the use of the Granularity-Related Inconsistency of Means – GRIM test as an AI tool for the detection of statistical errors [14].

A 2026 evaluation of an LLM agent across five stages of clinical data analysis found that it performed reliably in generating research questions and statistical analysis plans, but systematic errors emerged during analysis execution. These findings highlight the need for expert verification before AI-generated results are reported [15].

That being said, human oversight is always necessary to review research results and ensure reliability. Therefore, only people can make the final decision regarding the obtained results of the study, either independently or with the use of AI tools, which significant

Correspondence to: Nemanja Rančić, University of Defence, Faculty of Medicine of the Military Medical Academy, Crnotravska 17, 11 000, Belgrade, Serbia. E-mail: nece84@hotmail.com

ly facilitate and accelerate this process. Future comparative studies are needed to evaluate data analyzed by AI programs *versus* data analyzed by humans, in order to identify the advantages and disadvantages of each approach to making final decisions. This would likely significantly improve the work of AI software both for analyzing raw data and for controlling already processed data.

Nemanja Rančić[*†], Filip Pilipović[‡], Milica Djurdjević[*]

[*]University of Defence, Faculty of Medicine of the Military Medical Academy, Belgrade, Serbia; [†]Military Medical Academy, Institute of Radiology, Belgrade, Serbia; [‡]Institute for Orthopaedic Surgery "Banjica", Belgrade, Serbia

R E F E R E N C E S

1. *Lee P, Bubeck S, Petro J.* Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. N Engl J Med 2023; 388(13): 1233–9. DOI: 10.1056/NEJMsr2214184.
2. *Zhang T, Jin CJ, Song Y, Li D.* Are Neural Networks Better than Machine Learning? A Comparative Study for Travel Mode Predictions. Systems 2025; 13(12): 1099. DOI: 10.3390/systems13121099.
3. *Siafakas N, Vasarmidi E.* Risks of Artificial Intelligence (AI) in Medicine. Pneumon 2024; 37(3): 40. DOI: 10.18332/pne/191736.
4. *Karan U, Elbek O.* Artificial Intelligence in Medicine. Thorac Res Pract 2025; 26(6): 366–73. DOI: 10.4274/ThoracResPract.2025.2025-1-9.
5. *Wah JNK.* The rise of robotics and AI-assisted surgery in modern healthcare. J Robot Surg 2025; 19(1): 311. DOI: 10.1007/s11701-025-02485-0.
6. *Elkohail A, Khalifa AM, Soffar A, Omar I, Abdulaziz M, Elsaket A,* et al. The Role of Robotic-Assisted Surgery in Orthopedic Practice: A Comprehensive Review of Opportunities, Challenges, and Future Directions. Cureus 2025; 17(9): e92337. DOI: 10.7759/cureus.92337.
7. *Alnaimat F, AlSamhori ARF, El Sharu H, Othman L, Oralbek A, Zimba O.* Artificial Intelligence in Detecting Statistical Errors: Implications for Authors, Reviewers, and Editors. J Korean Med Sci 2025; 40(49): e342. DOI: 10.3346/jkms.2025.40.e342.
8. *Di Bella E, Preti S.* Artificial Intelligence in Statistics Education: Leveraging LLMs for Analysis and Learning. Trends High Educ 2026; 5(2): 39. DOI: 10.3390/higheredu5020039.
9. *Schwarz J.* The use of generative AI in statistical data analysis and its impact on teaching statistics at universities of applied sciences. Teach Stat 2025: 47: 118–28. DOI: 10.1111/test.12398.
10. *Nikolić B, Popović T.* Hypothesis testing and statistical test selection: Fundamentals of statistics in clinical studies - part II. Med Pregl 2024; 77(1–2): 49–54. DOI: 10.2298/MPNS2402049N
11. *Komić J, Bokonjić D, Rančić N.* Selected methods of statistical analysis for biomedical research. Banja Luka: University Medical Faculty; 2018. p. 174. (Serbian)
12. *Gibney E.* AI tools are spotting errors in research papers: inside a growing movement. Nature 2025. DOI: 10.1038/d41586-025-00648-5.
13. *Pellegrina D, Helmy M.* AI for scientific integrity: detecting ethical breaches, errors, and misconduct in manuscripts. Front Artif Intell 2025; 8: 1644098. DOI: 10.3389/frai.2025.1644098. Erratum in: Front Artif Intell 2025; 8: 1699373. DOI: 10.3389/frai.2025.1699373.
14. *Brown NJL, Heathers JAJ.* The GRIM Test: a simple technique detects numerous anomalies in the reporting of results in psychology. Soc Psychol Personal Sci 2017; 8(4): 363–9. DOI: 10.1177/1948550616673876.
15. *Wu Y, Fu DJ, Zhou Y, Wagner SK, Keane PA.* Performance, Failures, and Oversight of a Large Language Model Agent for Clinical Data Analysis: Evaluation Study. J Med Internet Res 2026; 28: e99597. DOI: 10.2196/99597.

Received on October 7, 2026
Accepted on October 21, 2026
Online First November 2026