

Editorial

Mario Plebani*

ChatGPT: Angel or Demond? Critical thinking is still needed

<https://doi.org/10.1515/cclm-2023-0387>

The Chatbot Generative Pre-trained Transformer (ChatGPT), developed by OpenAI, a type of Artificial Intelligence (AI) software designed to simulate conversations with humans, works through algorithms programmed to interpret natural language inputs, and provides appropriate pre-written or newly generated responses. The release of ChatGPT in November 2022 prompted immediate enthusiasm thanks to its multiple potential uses, but concerns were also expressed about its potential misuse, since the language model could be used to cheat on homework assignments, write essays, take examinations, and write scientific manuscripts for publication. In particular, there is heated debate amongst Journal editors, researchers, and publishers concerning the role of AI tools, such as ChatGPT, in publications. Artificial intelligence (AI) technologies designed to help authors improve on the preparation and the quality of their manuscripts and published articles are, in fact, rapidly increasing in number and sophistication. These include tools to assist with writing, grammar, language, references, statistical analysis, and reporting standards. Editors and publishers also use AI-assisted tools for myriad purposes, including screening submissions for problems (e.g., plagiarism, image manipulation, ethical issues), triage submissions, validation of references, editing, and providing code content for publication in different media as well as facilitating post-publication searches and discoverability [1]. After the publication of articles that included ChatGPT as a byline author, Nature, Science, JAMA and other journals defined a new policy to guide the use of large-scale artificial intelligence models in scientific publications [1–3]. The policy prohibits the naming of such tools as a “credited author on a research paper” because “attribution of authorship carries with it accountability for the work, and AI tools cannot take such responsibility” [2]. In addition, the policy also advises

researchers and authors who use these tools to document this use in the Methods or Acknowledgment sections of manuscripts. At a recently organized meeting of the Associate Editors Board of the *Journal*, we therefore decided to adopt a similar policy, which prohibits the addition of ChatGPT in the authorship of an article, and to request the authors to acknowledge its use in preparing the manuscript, as well as in the statistical elaboration of data, and preparation of figures and tables.

Appearing in this current issue of the *Journal* is an interesting article by Cadamuro and Colleagues, written on behalf of the Working Group on Artificial Intelligence (WG-AI) of the European Federation of Laboratory Medicine (EFLM). Entitled “potential and pitfalls of ChatGPT and natural-languages artificial intelligence models for the understanding of laboratory medicine test results” [4], it deserves merit and should give rise to further considerations. It should be noted that the paper contains an erroneous citation of an article in which ChatGPT was initially included in the authorship [5]; the authors, in fact, have subsequently published an errata corripse, stating that “shortly after initial publication of this article, the authorship and Acknowledgements were updated in line with Springer Nature authorship policies. Large Language Models (LLM), such as ChatGPT, do not currently satisfy our authorship criteria. An attribution of authorship carries with it accountability for the work, which cannot be effectively applied to LLMs,” [6]. This, in turn, may further endorse the decision to update the Information for Authors of the *Journal* to recognize the need for a policy that addresses the issue of the use of AI, particularly ChatGPT, in scientific publications.

Importantly, the paper by Cadamuro and Coll. provides the opportunity to further discuss the use of ChatGPT in laboratory medicine, particularly in improving the post-analytical phase, namely the interpretation of laboratory results. The main assumption in the study design was to “consider the use of a patient receiving his/her laboratory results after a routine check-up at his/her general practitioner and not a specific diagnostic question...” and “under the assumption that a patient has not yet discussed the results with his/her GP” [4]. Thereafter, ten plausible fictional clinical cases are described by the authors and the data from

*Corresponding author: Mario Plebani, Honorary Professor, Clinical Biochemistry and Clinical Molecular Biology, University of Padova, Padova, Italy; and Adjunct Professor, Department of Pathology, Medical Branch, University of Texas, Galveston, USA, E-mail: mario.plebani@unipd.it.
<https://orcid.org/0000-0002-0270-1711>

a series of ten common laboratory tests and some additional laboratory parameters provided to ChatGPT, along with the reference intervals, the age and the biological sex of the patient. My first observation concerns the choice of providing ChatGPT with only one of the possible comparators (i.e., the reference range), while evidence has been collected to demonstrate that, at least for most of the laboratory parameters used in this study, including glucose, total cholesterol, HDL, LDL and PSA, reference ranges should be avoided as they represent a confounding information, while the clinical decision and/or target levels should be reported. My second concern regards the concept of laboratory testing performed for “check-up”, which is known to be not only a “non-evidence-based” approach, but also an inappropriate and costly approach that should be discouraged. My third, and possibly most important concern, regards a statement made by the authors on answering one of the referees’ comments, in which they suggest *“interpreting laboratory test results without any additional clinical information is the most common type of standard scenario”*. This sentence is not only misleading, but also fails to acknowledge the work undertaken in the last few decades, in particular during the COVID-19 pandemic, to correctly address the issue of laboratory results’ interpretation. In a paper dealing with lessons from the COVID-19 pandemic, I wrote that “regarding the post-analytical phase, particular concern has finally been raised concerning the need to interpret any laboratory (and diagnostic) test result in the context of the pre-test probability of disease” [7]. The rediscovery of this seminal concept, well known by the laboratory science community, comes from the clinical side [8], and we can no longer perpetuate a mistake in considering the laboratory result “standalone information”. It is increasingly evident that laboratory data provide information of fundamental importance only if integrated in the context of an appropriate request and interpreted on the basis of other clinical information, including the pre-test probability. It appears absurd to expect an AI tool to make a better interpretation of laboratory results than that made by a human, hopefully a well-trained physician, if the data provided are limited. In addition, the authors’ approach should be revised on the basis of some fundamental ethical issues concerning the relationship between laboratory professionals and clinicians, and this can no longer be ignored. On the other hand, the use of ChatGPT or any other AI tool should be allowed in the context of self-testing and direct-to-consumer laboratory testing (DTCT). In this case, the patient without any

previous request by his/her GP or other physician must responsibly use the laboratory result to understand and predict the state of his/her health. This approach, in the context of so-called “patient empowerment”, has been claimed to “generate beneficial and actionable medical information” [9]. However, as underlined by some authors “both inappropriate testing, wrong interpretation of test results, and fraudulent invoicing of DTCT will have an impact on the public trust of IVD testing, irrespective of the setting in which this testing is performed, and will therefore have major implications on the future work of specialists in laboratory medicine” [10]. In conclusion, the paper by Cadamuro and colleagues is welcome, as it should force us to reconsider the issue of the true nature of laboratory testing, of fundamental ethical questions as well as the need to improve the interpretation of laboratory results, an issue which deserves further consideration and research.

References

1. Flanagan A, Bibbins-Domingo K, Berkswits M, Christiansen SL. Nonhuman “authors” and implications for the integrity of scientific publication and medical knowledge. *JAMA* 2023;329:637–9.
2. Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature* 2023;613:612.
3. Thorp HH. ChatGPT is fun, but not an author. *Science* 2023;379:313.
4. Cadamuro J, Cabitza F, Debeljak Z, De Bruyne S, Frans G, Perez SM, et al. Potentials and pitfalls of ChatGPT and natural-language artificial intelligence models for the understanding of laboratory medicine test results. An assessment by the European federation of clinical chemistry and laboratory medicine (EFLM) working group on artificial intelligence (WG-AI). *Clin Chem Lab Med* 2023;61: 1158–66.
5. Salvagno M, ChatGPT, Taccone FS, Gerli AG. Can artificial intelligence help for scientific writing? *Crit Care* 2023;27:75.
6. Salvagno M, Taccone FS, Gerli AG. Correction to: can artificial intelligence help for scientific writing? *Crit Care* 2023;27:99. Erratum *cor/for: Crit Care* 2023 Feb 25;27:75.
7. Plebani M. Laboratory medicine in the COVID-19 era: six lessons for the future. *Clin Chem Lab Med* 2021;59:1035–45.
8. Watson J, Whiting PF, Brush JE. Interpreting a Covid-19 test result. *BMJ* 2020;369:m1808.
9. Gill EL, Master SR. Big data everywhere: the impact of data disjunction in the direct-to-consumer testing model. *Clin Lab Med* 2020;40:51–9.
10. Orth M, Vollebregt E, Trenti T, Shih P, Tollanes M, Sandberg S. Direct-to-consumer laboratory testing (DTCT): challenges and implications for specialists in laboratory medicine. *Clin Chem Lab Med* 2022;61: 696–702.