

Review

Not peer-reviewed version

Chat GPT in Diagnostic Human Pathology: Will It Really Be Useful to the Pathologist?

[Gerardo Cazzato](#) ^{*}, [Marialessandra Capuzzolo](#), [Paola Parente](#), [Francesca Arezzo](#), [Vera Loizzi](#), [Andrea Marzullo](#), [Gennaro Cormio](#), [Giuseppe Ingravallo](#)

Posted Date: 29 August 2023

doi: [10.20944/preprints202308.1990.v1](https://doi.org/10.20944/preprints202308.1990.v1)

Keywords: ChatGPT; chatbot; Artificial Intelligence; AI; Pathology; Histology



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

Chat GPT in Diagnostic Human Pathology: Will It Really Be Useful to the Pathologist?

Gerardo Cazzato ¹, Marialessandra Capuzzolo ¹, Paola Parente ², Francesca Arezzo ³, Vera Loizzi ³, Andrea Marzullo ¹, Gennaro Cormio ³ and Giuseppe Ingravallo ¹

¹ Section of Molecular Pathology, Department of Precision and Regenerative Medicine and Ionian Area (DiMePRe-J), University of Bari "Aldo Moro", Bari, 70124, Italy. gerardo.cazzato@uniba.it (G.C.); m.capuzzolo@studenti.uniba.it (M.A.C.); andrea.marzullo@uniba.it (A.M.); giuseppe.ingravallo@uniba.it (G.I.)

² Pathology Unit, Fondazione IRCCS Ospedale Casa Sollievo della Sofferenza, San Giovanni Rotondo (FG) Italy.

³ IRCCS Istituto Tumori "Giovanni Paolo II", 70124 Bari, Italy. gennaro.cormio@uniba.it (G.C.)

* Correspondence: gerardo.cazzato@uniba.it (G.C.); Tel.: +39-3405203641

Abstract: The advent of Artificial Intelligence (AI) has in just a few years invested multiple areas of knowledge, also affecting the medical-scientific sector. An increasing number of AI-based applications have been developed, among which conversational AI has emerged. Among these, ChatGPT has risen to the headlines, scientific and otherwise, for its distinct propensity to simulate a 'real' discussion with its interlocutor, based on appropriate prompts. Although several clinical studies using ChatGPT have already been published in the literature, very little has yet been written about its potential application in human pathology. We conduct a systematic review following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, using PubMed and Scopus as databases, with the following keywords: ChatGPT OR Chat GPT, in combination with each of the following: Pathology, diagnostic pathology, anatomic pathology. A total of 90 records were initially identified in the literature search, of which 6 were duplicates. After screening for eligibility and inclusion criteria, only 5 publications were ultimately included. The majority of publications were original articles (n = 2), followed by case reports (n = 1), letter to the editor (n = 1) and review (n = 1). Although the premises are exciting and ChatGPT is able to co-advice the pathologist in providing large amounts of scientific data for use in routine microscopic diagnostic practice, there are many limitations that need to be addressed and resolved, with the caveat that an AI-driven system should always provide support and never a decision-making motive during the anatomo-pathological diagnostic process.

Keywords: ChatGPT; chatbot; artificial intelligence; AI; pathology; histology

1. Introduction

Artificial Intelligence (AI) has revolutionized the medical-scientific field in just a few years, allowing for significant changes and integration of diagnostic, therapeutic and patient care pathways [1]. Although at first it was mainly represented by the development of Machine Learning (ML) models [2], further advances such as Deep Learning (DL) with, among others, Convolutional Neural Networks (CNN) soon came to the fore [3]. A branch of AI includes conversational artificial intelligence, which has experienced unprecedented development in recent years, with numerous models and platforms developed to enable machines to understand and respond to natural language input [4]. In more detail, a chatbot is an item of software that simulates and develops human conversations (spoken or written), allowing users to interact with digital devices as though they were speaking with real people [5]. Chatbot might be as basic as a programme that responds to a single enquiry or as complex as a digital assistant that learns and develops as it gathers and elaborates information to provide more and higher levels of personalization [6]. The chatbots designed



specifically for activities (declarational) are '*single-purpose*' softwares that focus on carrying out a certain function; regulated responses to user requests are generated using Natural Language Process (NLP) and very little machine learning [7]. The interactions with these chatbots are quite particular and structured, and they are best suited for assistance and service functions like frequently asked and consolidated questions. Common questions can be managed by activity-specific chatbots, such as inquiries about working hours or straightforward transactions that don't involve many variables. Even while they employ NLP in a way that allows users to experiment with it easily, their capabilities are still somewhat limited. These are the most popular chatbots right now [7]. Virtual assistants, also known as digital assistants or data-driven predictive (conversational) chatbots, are significantly more advanced, interactive, and customized than task-specific chatbots. These chatbots use ML, NLP and context awareness to learn. They employ data analysis and predictive intelligence to offer customization based on user profiles and past user behaviour. Digital assistants can gradually learn a user's preferences, make suggestions, and even foresee needs. They can start talks in addition to monitoring data and rules. Predictive chatbots that focus on the needs of the user and are data-driven include Apple's Siri and Amazon's Alexa [8].

A clear example of such an approach is Chat GPT, an acronym for Generative Pretrained Transformer, a powerful and versatile natural language processing (or Natural Language Processing) tool that uses advanced machine learning algorithms to generate human-like responses within a conversation (<https://chat.openai.com>). Released on 30 November 2022 by OpenAI, Chat GPT was trained until the end of 2021 on more than 300 billion words, with the ability to respond on a huge variety of topics and with the ability to learn from its human interlocutor [9]. Released on 30 November 2022 by OpenAI (San Francisco, CA), Chat GPT was trained until the end of 2021 on more than 300 billion words, with the ability to respond on a huge variety of topics and with the ability to learn from its human interlocutor [9]. In the first few months after its official launch, many papers were published in the purely informatic field, but, as the weeks went by, the medical-scientific field was also interested, with a particular interest in education, research and simulation of clinical pictures of patients, as well as applications in hygiene and public health, clinical medicine, oncology and surgery. In this narrative review, however, we focus on the potential use of Chat GPT in Pathological Anatomy, discuss the fields of application studied so far and try to outline future perspectives, with particular regard to present limitations.

2. Materials and Methods

A systematic review was elaborated following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, using PubMed and Scopus databases until 07/31/2023 with the following terms: ChatGPT OR Chat GPT, in combination with each of the following: Pathology, diagnostic pathology, anatomic pathology. Only articles in English were recorded. Review articles, meta-analyses, observational studies, case reports, survey snapshot studies, letters to the editor, and comments to the letters were all included. Other potentially relevant articles were identified by manually checking the references of the included literature. The articles all had to meet the following inclusion criteria: 1) Covering pathological anatomy topics in light of the use of Chat GPT, with the opportunity to discuss strengths and/or limitations.

An independent extraction of articles was performed by two investigators (G.C. and M.A.C.) according to the inclusion criteria. Disagreement was resolved by discussion between the two review authors.

3. Results

A total of 90 records were initially identified in the literature search, of which 6 were duplicates. After screening for eligibility and inclusion criteria, only 5 publications were ultimately included (Figure 1). The majority of publications were original articles (n = 2), followed by case reports (n = 1), letter to the editor (n = 1) and review (n = 1).

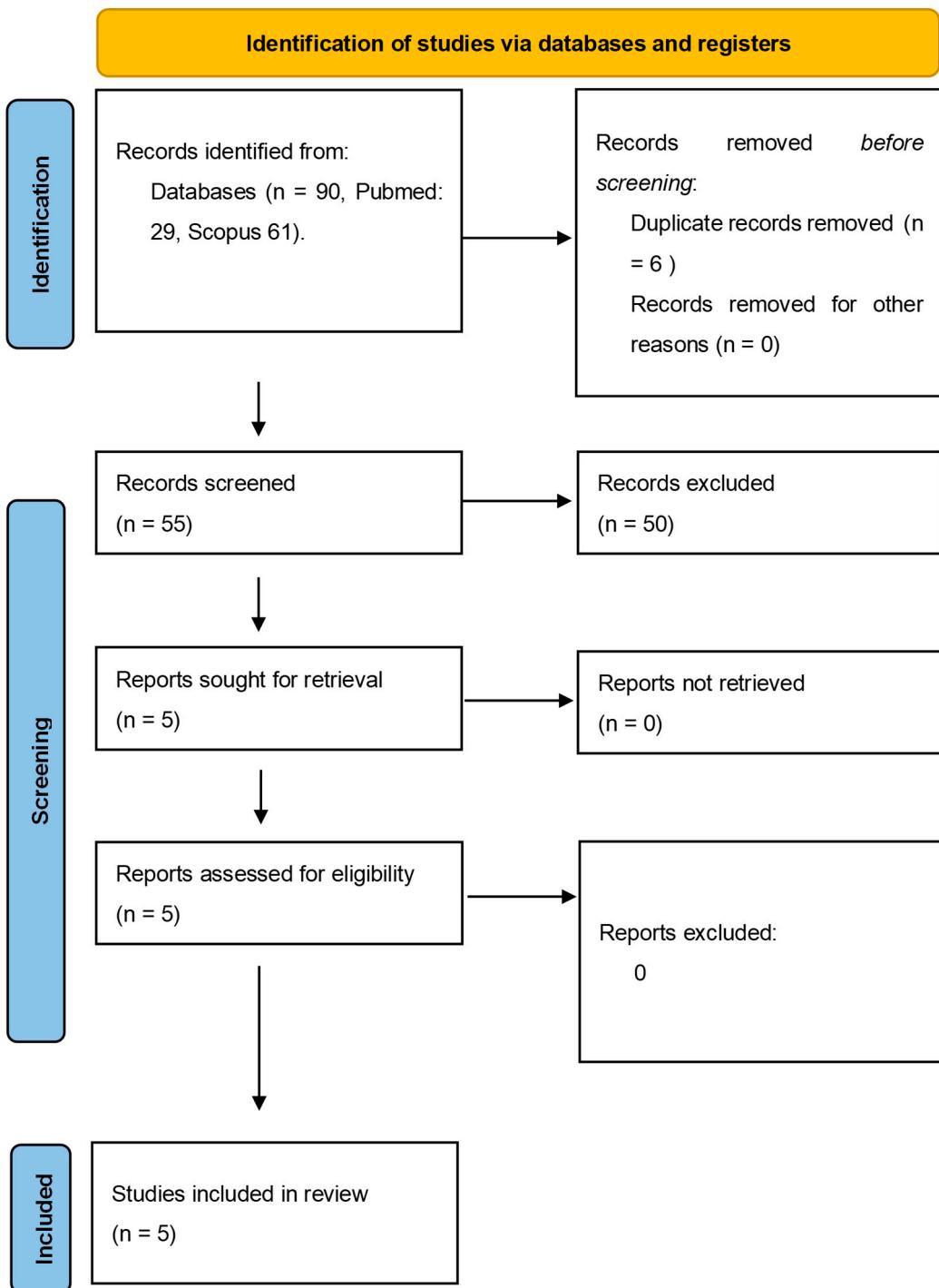


Figure 1. PRISMA flow-chart using in this review.

Figure 1 summarize the review process following PRISMA guidelines.

4. Discussion

The application of AI to medicine has significantly co-assisted the physician's therapeutic decision-making processes, not replacing, but complementing and enhancing the indispensable figure of the human [10]. The advent of Chat GPT has further enabled a breakthrough in Large Language Models (LLM) that enable the simulation of a real conversation on a wide variety of topics, including medical-scientific notions [11]. If in the early months the field of application was mostly restricted to clinical medicine, in recent times a number of papers have studied, tested and

commented on the applicability of Chat GPT in the field of human pathology, making it possible to outline its real usefulness and current limitations.

The article by Sinha et al. [12] describes a study conducted on ChatGPT's ability to resolve complex rationality problems in the area of human pathology. Based on the clear finding that AI is used to analyze medical images, such as histopathologic slides, in order to identify and diagnose diseases with high precision, the authors take a cautious approach to the fact that NLP algorithms are used to analyze the relationships between pathologies, extract relevant information, and aid in disease diagnosis. The goal of the study was to assess ChatGPT's ability to address high-level rational questions in the field of pathology. 100 questions that were randomly chosen from a bank of inquiries regarding diseases and divided into 11 different systems of pathology were used. Experts have evaluated the responses provided by ChatGPT using both a scale of 0 to 5 and the taxonomy SOLE to assess the depth of understanding demonstrated in the responses.

The outcomes have shown that the responses provided by ChatGPT achieve a reasonable level of rationality, with a score similar to four on five. This means that the AI is capable of correctly responding to high-level inquiries requiring in-depth knowledge of the subject. The report also highlights the limitations of AI in diagnosing disease. Although it is possible to recognize schemes and categorize data, a true understanding of the underlying significance and context of the information is lacking. L'AI is unable to make logical judgements or evaluative decisions because it lacks the ability to comprehend personal values and judgements. Therefore, it is suggested that careful consideration should be given to the use of AI in medical education, with the goal of assisting human judgement rather than replacing it.

There are some limitations to the study, including the subjective nature of the evaluation procedure and the selection of particular questions from a single bank of data. The authors suggest that in order to obtain results that are more generally applicable, future studies may be conducted on a larger sample size and by a variety of institutions.

In the paper by Sorin V. et al [13], the authors discuss how ChatGPT 3.5 can also operate within the molecular tumour board, not only starting from histopathological/diagnostic data, but also integrating other key components such as genetic, molecular, response and/or prediction of treatment response and prognosis data. Ten consecutive cases of women with breast cancer were considered and an attempt was made to assess how consistent the recommendations provided by the chatbot were with those of the tumour board. The results showed that ChatGPT's clinical recommendations were in line with those of the oncology committee in 70% of the cases, with concise clinical case summaries and explained and reasoned conclusions. However, the lowest scores (which were given by the second reviewer) were for the clinical recommendations of the chatbot, suggesting that deciding on clinical treatment from pathological/molecular data is highly challenging, requiring medical understanding and experience in the field. Furthermore, it was curious to note that ChatGPT never mentioned the role of medical radiologists, suggesting the importance (and consequent risk of bias) that incomplete training may influence the performance, and thus the responses, of the chatbot.

In another recent paper by Naik H.R. et al [14], the authors described the case of a 58-year-old woman with bilateral synchronous breast cancer (s-BBC) who underwent bilateral mastectomy, sentinel lymph node biopsy (BLS), axillary lymphadenectomy with adjuvant radiotherapy and chemo/hormonotherapy. The particularity of the paper was related to the so-called 'hallucination phenomenon', i.e. a clear and confident response from ChatGPT but which is not real. One of the authors of the paper (Dr. Gurda), when asking the chatbot about s-BBC, noted that although the answer was plausible, the reference provided did not exist, although there were articles with similar information and authors.

This aspect is well addressed by the paper by Metze K. et al [15] who conducted a study to assess the ability of ChatGPT to contribute to a review on Chagas disease, focusing on the role of individual researchers. Therefore, 50 names of researchers with at least 4 publications on Chagas disease were selected from Clarivate's Web of Sciences (WoS) database and for each researcher, the chatbot was asked to provide conceptual contributions related to the study of the disease. The answers were checked by two observers against the literature, and incorrect information was removed. The

percentage of correct words in the text generated by ChatGPT was calculated and the literature references were classified into three categories: completely correct, minor errors and major errors.

The results showed that the average percentage of correct words in the text generated by ChatGPT was 59.4%, but the variation was wide, ranging from 10.0% to 100.0%. A positive correlation was observed between the percentage of correct words and the number of indexed publications of each author of interest, as well as with the number of citations and the author's H-index. However, the percentage of correct references was very low, averaging 7.07%, and both minor and major errors were found in the references.

In conclusion, the results of this study suggest that ChatGPT is still not a reliable source of literature review, especially in more specific areas with a relatively low number of publications, as there are still accuracy and misinformation issues to be addressed, especially in the field of medicine.

Yamin Ma [16], in a paper of July 2023, discusses the application of ChatGPT in the context of gastrointestinal pathology, hypothesizing three possible applications for ChatGPT: 1) Ability to summarize patient records: ChatGPT could be integrated into the patient table to summarize patients' previous clinical information, helping pathologists better understand patients' current health status and saving time before case reviews.

2) Incorporation into digital pathology: ChatGPT could improve the interpretation of computer-aided diagnosis (CAD) systems in gastrointestinal pathology. It would enable pathologists to ask specific questions on digitized images and obtain knowledge-based answers associated with diagnostic criteria and differential diagnosis.

3) Role in education and research: ChatGPT could be used for health education, offering scientific explanations associated with medical terms in pathology. However, attention should be paid to the quality of the training data to avoid biased content and inaccurate information. The use of ChatGPT in research also requires caution as it may be insufficient or misleading.

Finally, the paper emphasizes that while recognizing the potential of ChatGPT, it is important to proceed with caution when using artificial intelligence-based technologies such as ChatGPT in gastrointestinal pathology. The aim should be to integrate such language models in a regulated and appropriate manner, exploiting their advantages to improve the quality of healthcare without replacing human expertise and without ignoring expert consultation in particular cases.

From what has been discussed so far and bearing in mind a paper published five days ago [17], it would appear that at present, the use of ChatGPT in pathology is still in its early stages. In particular, with regard to ChatGPT version 3.5, it seems clear that the amount of data on which the algorithm has been trained plays a key role in its ability to provide correct answers to certain prompts. In particular, several papers have warned of the risk of possible bias and transparency issues [18,19] and of damage resulting from inaccurate or outright incorrect content [20–23]. One of the most problematic phenomena is hallucination, as ChatGPT seems, at present, to produce correct scientific content but not to direct the content itself to a real source/reference. Therefore, its use in pathology and, more generally, in scientific research must necessarily take these limitations into account [24].

Moreover, projecting to a future in which systems may become more performant, the integration with clinical, genetic, anamnestic, morphological and immunohistochemical data, which have always characterised the pathologist's life, will always have to be screened by medical experience and knowledge, areas in which ChatGPT struggles most.

Author Contributions: Conceptualization, G.C. and M.A.C.; methodology, G.C.; software, G.C.; validation, G.C., M.A.C. and G.I.; formal analysis, P.P.; investigation, G.C.; resources, G.C.; data curation, G.C.; writing—original draft preparation, G.C.; writing—review and editing, G.C.; visualization, G.I.; supervision, P.P. and G.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Data Availability Statement: Not applicable

Acknowledgments: Not applicable

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Xiang, Y.; Zhao, L.; Liu, Z.; Wu, X.; Chen, J.; Long, E.; Lin, D.; Zhu, Y.; Chen, C.; Lin, Z.; Lin, H. Implementation of artificial intelligence in medicine: Status analysis and development suggestions. *Artif Intell Med.* **2020** Jan;102:101780. doi: 10.1016/j.artmed.2019.101780. Epub 2019 Dec 18. PMID: 31980086.
2. Haug, C.J.; Drazen, J.M. Artificial Intelligence and Machine Learning in Clinical Medicine, 2023. *N Engl J Med.* **2023** Mar 30;388(13):1201-1208. doi: 10.1056/NEJMra2302038. PMID: 36988595.
3. Venerito, V.; Angelini, O.; Cazzato, G.; Lopalco, G.; Maiorano, E.; Cimmino, A.; Iannone, F. A convolutional neural network with transfer learning for automatic discrimination between low and high-grade synovitis: a pilot study. *Intern Emerg Med.* **2021** Sep;16(6):1457-1465. doi: 10.1007/s11739-020-02583-x. Epub 2021 Jan 2. PMID: 33387201.
4. Eysenbach, G. The Role of ChatGPT, Generative Language Models, and Artificial Intelligence in Medical Education: A Conversation With ChatGPT and a Call for Papers. *JMIR Med Educ.* **2023** Mar 6;9:e46885. doi: 10.2196/46885. PMID: 36863937; PMCID: PMC10028514.
5. Bozic, J.; Tazl, O.A.; Wotawa, F. Chatbot Testing Using AI Planning, 2019 IEEE International Conference On Artificial Intelligence Testing (AITest), Newark, CA, USA, **2019**, pp. 37-44, doi: 10.1109/AITest.2019.00-10.
6. Adamopoulou, E.; Moussiades, L. An Overview of Chatbot Technology. *Artificial Intelligence Applications and Innovations.* **2020** May;6:584:373-83. doi: 10.1007/978-3-030-49186-4_31. PMCID: PMC7256567.
7. Chinedu Wilfred Okonkwo, Abejide Ade-Ibijola, Chatbots applications in education: A systematic review, *Computers and Education: Artificial Intelligence*, Volume 2, 2021,100033.
8. Caldarini, G.; Jaf, S.; McGarry, K. A Literature Survey of Recent Advances in Chatbots. *Information.* **2022**, 13, 41. <https://doi.org/10.3390/info13010041>.
9. Panch, T.; Pearson-Stuttard, J.; Greaves, F.; Atun, R. Artificial intelligence: opportunities and risks for public health. *Lancet Digit Health.* **2019** May;1(1):e13-e14. doi: 10.1016/S2589-7500(19)30002-0. Epub 2019 May 2. Erratum in: *Lancet Digit Health.* 2019 Jul;1(3):e113. PMID: 33323236.
10. Khan, R.A.; Jawaid, M.; Khan, A.R.; Sajjad, M. ChatGPT - Reshaping medical education and clinical management. *Pak J Med Sci.* **2023** Mar-Apr;39(2):605-607. doi: 10.12669/pjms.39.2.7653. PMID: 36950398; PMCID: PMC10025693.
11. Mago, J.; Sharma, M. The Potential Usefulness of ChatGPT in Oral and Maxillofacial Radiology. *Cureus.* **2023** Jul 19;15(7):e42133. doi: 10.7759/cureus.42133. PMID: 37476297; PMCID: PMC10355343.
12. Sinha, R.K.; Deb Roy, A.; Kumar, N.; Mondal, H. Applicability of ChatGPT in Assisting to Solve Higher Order Problems in Pathology. *Cureus.* **2023** Feb 20;15(2):e35237. doi: 10.7759/cureus.35237. PMID: 36968864; PMCID: PMC10033699.
13. Sorin, V.; Klang, E.; Sklair-Levy, M.; Cohen, I.; Zippel, D.B.; Balint Lahat, N.; Konen, E.; Barash, Y. Large language model (ChatGPT) as a support tool for breast tumor board. *NPJ Breast Cancer.* **2023** May 30;9(1):44. doi: 10.1038/s41523-023-00557-8. PMID: 37253791; PMCID: PMC10229606.
14. Naik, H.R.; Prather, A.D.; Gurda, G.T. Synchronous Bilateral Breast Cancer: A Case Report Piloting and Evaluating the Implementation of the AI-Powered Large Language Model (LLM) ChatGPT. *Cureus.* **2023** Apr 14;15(4):e37587. doi: 10.7759/cureus.37587. PMID: 37193434; PMCID: PMC10183235.
15. Metze, K.; Morandin-Reis, R.C.; Lorand-Metze, I.; Florindo, J.B. The Amount of Errors in ChatGPT's Responses is Indirectly Correlated with the Number of Publications Related to the Topic Under Investigation. *Ann Biomed Eng.* **2023** Jul;51(7):1360-1361. doi: 10.1007/s10439-023-03205-1. Epub 2023 Apr 15. PMID: 37061596.
16. Yamin Ma. The potential application of ChatGPT in gastrointestinal pathology. *Gastroenterology & Endoscopy*, Volume 1, Issue 3, 2023, Pages 130-131.
17. Schukow, C.; Smith, S.C.; Landgrebe, E.; Parasuraman, S.; Folaranmi, O.O., Paner, G.P., Amin, M.B. Application of ChatGPT in Routine Diagnostic Pathology: Promises, Pitfalls, and Potential Future Directions. *Adv Anat Pathol.* **2023** Jul 27. doi: 10.1097/PAP.0000000000000406. Epub ahead of print. PMID: 37501529.
18. A. Holzinger, K. Keiblanger, P. Holub, K. Zatloukal, H. Müller. AI for life: trends in artificial intelligence for biotechnology. *N Biotech.* 74 (2023), pp. 16-24.
19. K. Jeblick, B. Schachtner, J. Dexl, et al. ChatGPT Makes Medicine Easy to Swallow: An Exploratory Case Study on Simplified Radiology Reports. *arXiv.* (2022) arXiv:2212.14882.
20. Ahn, C. Exploring ChatGPT for information of cardiopulmonary resuscitation. *Resuscitation.* **2023** Apr;185:109729. doi: 10.1016/j.resuscitation.2023.109729. Epub 2023 Feb 10. PMID: 36773836.
21. D'Amico, R.S.; White, T.G.; Shah, H.A.; Langer, D.J.; I Asked a ChatGPT to Write an Editorial About How We Can Incorporate Chatbots Into Neurosurgical Research and Patient Care.... *Neurosurgery.* **2023** Apr 1;92(4):663-664. doi: 10.1227/neu.0000000000002414. Epub 2023 Feb 9. PMID: 36757199.

22. Patel, S.B.; Lam, K. ChatGPT: the future of discharge summaries? *Lancet Digit Health.* **2023** Mar;5(3):e107-e108. doi: 10.1016/S2589-7500(23)00021-3. Epub 2023 Feb 6. PMID: 36754724.
23. Ali, S.R.; Dobbs, T.D.; Hutchings, H.A.; Whitaker, I.S. Using ChatGPT to write patient clinic letters. *Lancet Digit Health.* **2023** Apr;5(4):e179-e181. doi: 10.1016/S2589-7500(23)00048-1. Epub 2023 Mar 7. PMID: 36894409.
24. Goddard, J. Hallucinations in ChatGPT: A Cautionary Tale for Biomedical Researchers. *Am J Med.* **2023** Jun 25:S0002-9343(23)00401-1. doi: 10.1016/j.amjmed.2023.06.012. Epub ahead of print. PMID: 37369274.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.