







Review

Chat GPT in Diagnostic Human Pathology: Will It Be Useful to Pathologists? A Preliminary Review with ‘Query Session’ and Future Perspectives

Gerardo Cazzato ^{1,*} , Marialessandra Capuzzolo ¹, Paola Parente ² , Francesca Arezzo ³, Vera Loizzi ³ ,
Enrica Macorano ⁴ , Andrea Marzullo ¹, Gennaro Cormio ³  and Giuseppe Ingravalle ¹ 

- ¹ Section of Molecular Pathology, Department of Precision and Regenerative Medicine and Ionian Area (DiMePRE-J), University of Bari “Aldo Moro”, 70124 Bari, Italy; m.capuzzolo@studenti.uniba.it (M.C.); andrea.marzullo@uniba.it (A.M.); giuseppe.ingravalle@uniba.it (G.I.)
 - ² Pathology Unit, Fondazione IRCCS Ospedale Casa Solievo della Sofferenza, 71013 San Giovanni Rotondo, Italy; paolaparente77@gmail.com
 - ³ IRCCS Istituto Tumori “Giovanni Paolo II”, 70124 Bari, Italy; francesca.arezzi@uniba.it (F.A.); vera.loizzi@uniba.it (V.L.); gennaro.cormio@uniba.it (G.C.)
 - ⁴ Section of Legal Medicine, Interdisciplinary Department of Medicine, University of Bari “Aldo Moro”, 70124 Bari, Italy; enricamacorano@gmail.com
- * Correspondence: gerardo.cazzato@uniba.it; Tel.: +39-340-520-3641

Abstract: The advent of Artificial Intelligence (AI) has in just a few years supplied multiple areas of knowledge, including in the medical and scientific fields. An increasing number of AI-based applications have been developed, among which conversational AI has emerged. Regarding the latter, 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, Scopus and the Web of Science (WoS) as databases, with the following keywords: ChatGPT OR Chat GPT, in combination with each of the following: pathology, diagnostic pathology, anatomic pathology, before 31 July 2023. A total of 103 records were initially identified in the literature search, of which 19 were duplicates. After screening for eligibility and inclusion criteria, only five publications were ultimately included. The majority of publications were original articles ($n = 2$), followed by a case report ($n = 1$), letter to the editor ($n = 1$) and review ($n = 1$). Furthermore, we performed a ‘query session’ with ChatGPT regarding pathologies such as pigmented skin lesions, malignant melanoma and variants, Gleason’s score of prostate adenocarcinoma, differential diagnosis between germ cell tumors and high grade serous carcinoma of the ovary, pleural mesothelioma and pediatric diffuse midline glioma. Although the premises are exciting and ChatGPT is able to co-advise the pathologist in providing large amounts of scientific data for use in routine microscopic diagnostic practice, there are many limitations (such as data of training, amount of data available, ‘hallucination’ phenomena) 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 histopathological diagnostic process.

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



Citation: Cazzato, G.; Capuzzolo, M.; Parente, P.; Arezzo, F.; Loizzi, V.; Macorano, E.; Marzullo, A.; Cormio, G.; Ingravalle, G. Chat GPT in Diagnostic Human Pathology: Will It Be Useful to Pathologists? A Preliminary Review with ‘Query Session’ and Future Perspectives. *AI* **2023**, *4*, 1010–1022. <https://doi.org/10.3390/ai4040051>

Academic Editor: Tim Hulsen

Received: 9 October 2023

Revised: 30 October 2023

Accepted: 6 November 2023

Published: 22 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Artificial Intelligence (AI) has revolutionized medical and scientific fields in just a few years, allowing for significant changes and the 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 program 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 higher levels of personalization [6]. The chatbots designed specifically for activities (declarational) are ‘single-purpose’ software 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 do not 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 [7,8]. 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 behavior. 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 [9].

A clear example of such an approach is ChatGPT, an acronym for Generative Pretrained Transformer, which is a powerful and versatile NLP tool that uses advanced machine learning algorithms to generate human-like responses within a conversation (<https://chat.openai.com>, accessed on 1 July 2023). Released on 30 November 2022 by OpenAI, ChatGPT (version 3.5) was trained until the end of 2021 on more than 300 billion words, with the ability to respond to a huge variety of topics and with the ability to learn from its human interlocutor [10]. 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 and scientific fields were also interested, with a particular interest in the education, research and simulation of clinical pictures of patients, as well as applications in hygiene and public health, clinical medicine, oncology and surgery [11].

On the other hand, in the literature there is a paucity of information regarding the reliability of ChatGPT in assisting the routine activity of the pathologist [12]; among other papers, a recent manuscript by Schukow C. et al. [12,13] underlined the lack of studies that evaluate this relationship, focusing more on the three fundamental criteria on which a potential use of ChatGPT should be based: (1) a chatbot should have a strong performance; (2) an ideal chatbot should be freely accessible for public use; (3) it should be trained on known and recoverable data.

In this review paper, we will try to summarize the potential use of ChatGPT in pathological anatomy, discuss the fields of application studied so far, perform some ‘query sessions’ about pathological topics that could help the pathologist 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, Scopus and Web of Science (WoS) databases before 31 July 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 ChatGPT, with the opportunity to discuss strengths and/or limitations; (2) the articles had to necessarily relate ChatGPT to pathology. Exclusion criteria were articles that talked about ChatGPT in general or relating it to other aspects not pertaining to pathological anatomy.

An independent extraction of articles was performed by two investigators (G.C. and M.C.) according to the inclusion criteria, before 31 July 2023. Disagreement was resolved by discussion between the two review authors (Figure 1).

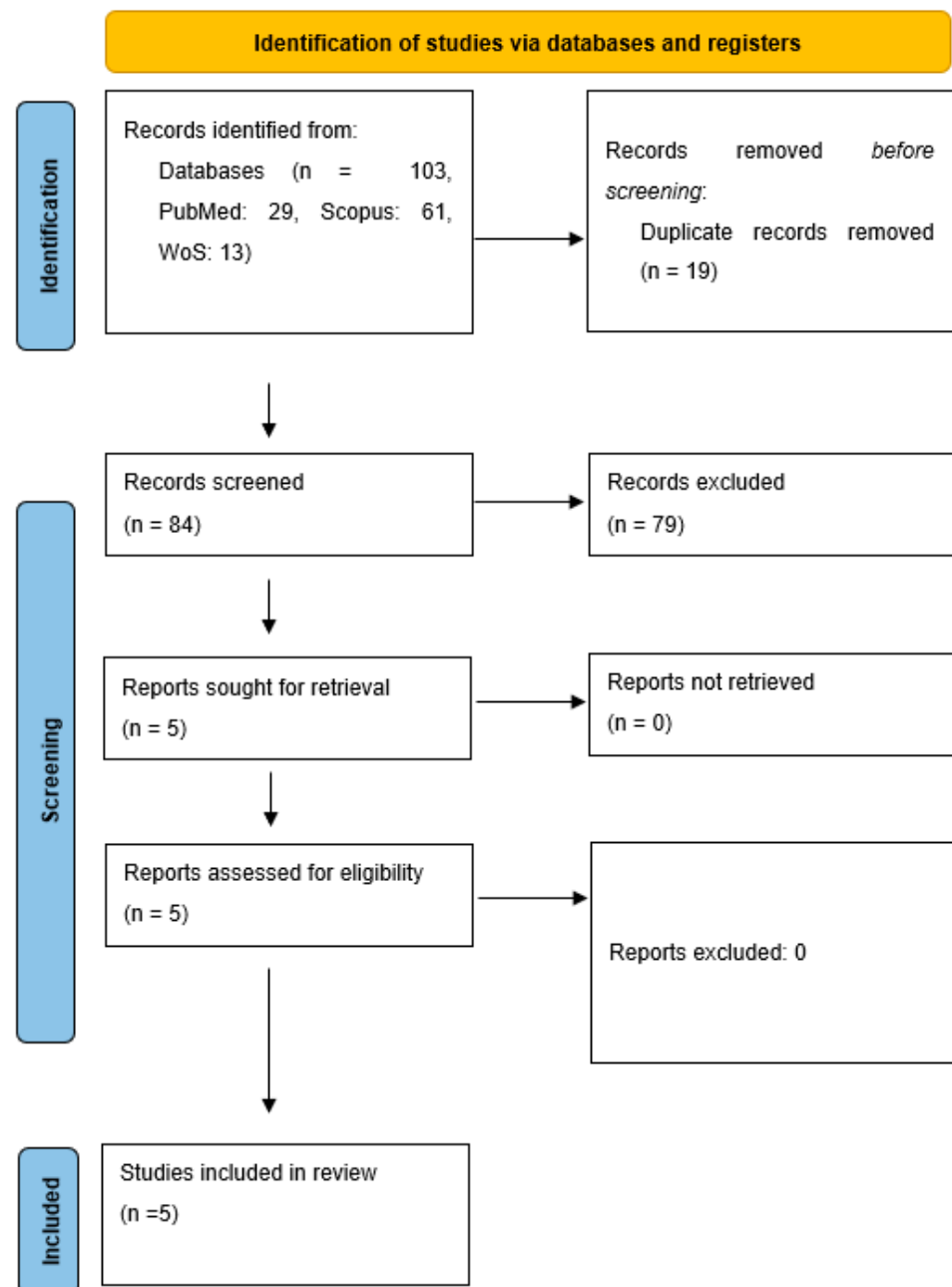


Figure 1. PRISMA flow-chart using in this review.

Furthermore, to explore the reliability of ChatGPT in potentially supporting the pathologist's routine diagnostic activity, we developed questions (G.C. and M.C.) that we asked ChatGPT regarding examples of diagnostic approaches in real-life patient cases. Specifically, the first author (G.C.) created a free account on the Open AI site and conducted the conversations with the chatbot. No plugins or other accessories were used in the platform, and all results presented were obtained with ChatGPT version 3.5 and the query sessions were performed between 1 October and 30 October 2023. Some examples of the conversations are summarized in Figures 2–6.

The image shows a screenshot of a ChatGPT interface. On the left, a vertical chat history pane shows the conversation context. The main chat area displays the user's query and the AI's response. The user asks for a differential diagnosis for a 32-year-old woman with a pigmented lesion on the thigh. The AI provides a systematic approach, including medical history, clinical examination, dermatoscopy, biopsy, differential diagnosis (listing various skin conditions), pathology evaluation, further tests, and treatment planning. It emphasizes the importance of consulting a healthcare professional.

User: Good evening ChatGPT?

ChatGPT: Good evening! How can I assist you today?

User: I want to see if you are useful to a pathologist

ChatGPT: I can certainly provide information and assistance related to pathology, medical terminology, and various aspects of the field. Whether you're looking for information on specific diseases, diagnostic techniques, medical research, or any other related topic, feel free to ask, and I'll do my best to help. Please keep in mind that I'm not a substitute for professional medical advice, diagnosis, or treatment, and you should always consult with a qualified healthcare provider or pathologist for any specific medical concerns or issues. How can I assist you with your questions related to pathology today?

User: ok, Could you make a differential diagnosis in a case of this type: woman, 32 years-old with a pigmented lesion on the thigh. What is the diagnostic process?

ChatGPT: Diagnosing a pigmented lesion on the thigh, also known as a skin lesion, requires a systematic approach that typically involves several steps. It's important to remember that I can't provide a definitive diagnosis, but I can outline a general process for the evaluation of such lesions and provide you with a list of possible differential diagnoses. For an accurate diagnosis, it's essential for the patient to consult a healthcare professional, such as a dermatologist. Here's a general diagnostic process and some potential differential diagnoses:

- 1. Medical History and Clinical Examination:**
 - The first step is to take a detailed medical history from the patient, which may include information about changes in the lesion, symptoms (such as pain, itching, or bleeding), family history of skin conditions, and any potential risk factors (e.g., sun exposure, tanning bed use).
 - A physical examination of the lesion is conducted to assess its size, shape, color, texture, and any other relevant features.
- 2. Dermatoscopy:**
 - Dermatoscopy is a non-invasive technique that involves using a handheld device to magnify the lesion and examine its surface patterns. It can provide valuable information to aid in diagnosis.
- 3. Biopsy:**
 - If the lesion is suspicious, a dermatologist may recommend a skin biopsy. This involves removing a small sample of tissue for examination under a microscope to determine its nature.
- 4. Differential Diagnosis:**
 - The differential diagnosis for a pigmented lesion on the thigh can include various possibilities, such as:
 - Melanocytic Nevus (moles)
 - Seborrheic Keratosis
 - Actinic Keratosis
 - Basal Cell Carcinoma
 - Squamous Cell Carcinoma
 - Melanoma
 - Dermatofibroma
 - Hemangioma
 - Pyogenic Granuloma
- 5. Pathology Evaluation:**
 - The biopsy sample is examined by a pathologist to determine the presence of abnormal cells, signs of malignancy, and other specific characteristics.
- 6. Further Tests:**
 - Depending on the initial findings, additional tests such as immunohistochemistry, molecular testing, or imaging studies may be recommended to confirm the diagnosis or assess the extent of disease.
- 7. Treatment Planning:**
 - Treatment options and recommendations will depend on the specific diagnosis. It could range from observation for benign lesions to various treatment modalities for malignancies.

It's essential for the patient to consult a healthcare professional, preferably a dermatologist, to determine the exact nature of the pigmented lesion on the thigh and establish the appropriate course of action. Early detection and intervention are crucial for skin health, especially when considering the possibility of skin cancer.

Figure 2. The first author (G.C.) posted a request to ChatGPT: “Could you make a differential diagnosis in a case of this type: woman, 32 years-old with a pigmented lesion on the thigh. What is the diagnostic process?” ChatGPT replied, “Diagnosing a pigmented lesion on the thigh, also known as a skin lesion, requires a systematic approach that typically involves several steps. It is important to remember that I cannot provide a definitive diagnosis, but I can outline a general process for the evaluation of such lesions and provide a list of possible differential diagnoses. For an accurate diagnosis, it is essential for the patient to consult a healthcare professional, such as a dermatologist. Here is a general diagnostic process and some potential differential diagnoses [. . .]” (see Figure 2 for the rest). The version of ChatGPT used was dated 25 September 2023, and the average time taken from request entry was about 23.43 s.

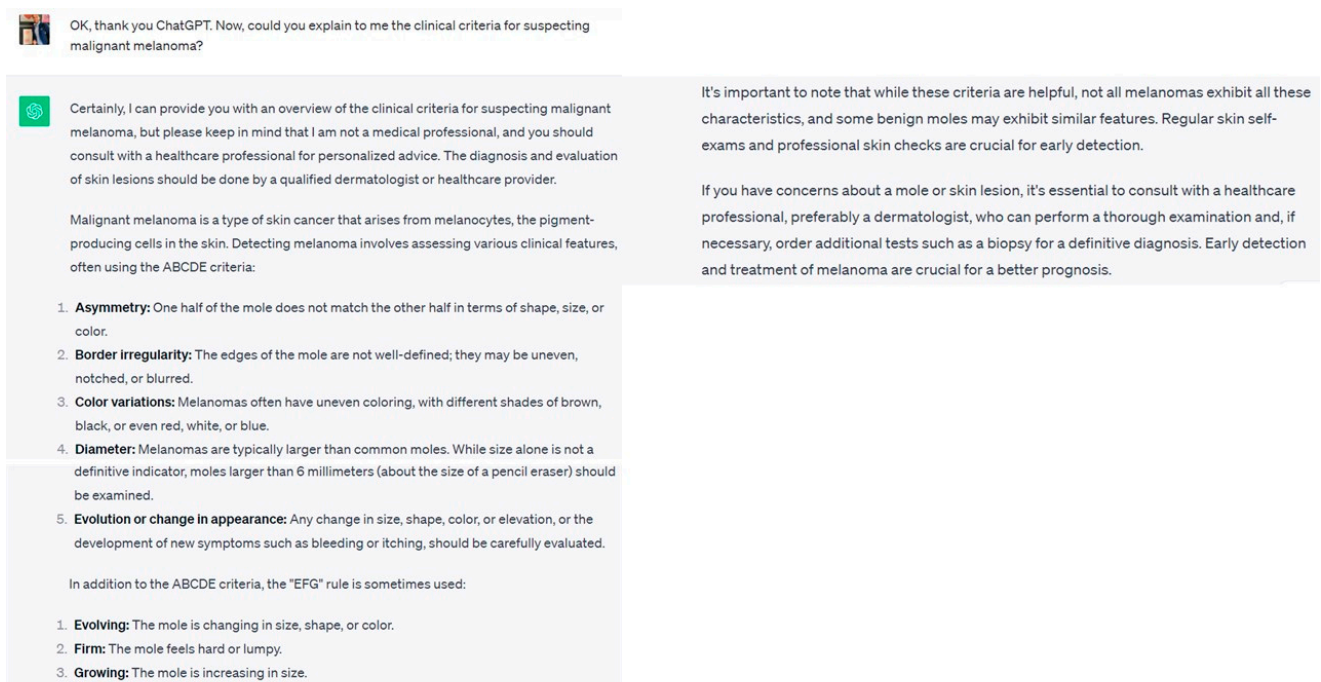


Figure 3. The first author (G.C.) followed up the initial enquiry with another question concerning clinical features useful for suspecting malignant melanoma. ChatGPT provided the ABCDE criteria together with other parameters (detectability, hardness/compactness and lesion growth). The version used was always that of 25 September 2023. ChatGPT had an average response time of 44.32 s.

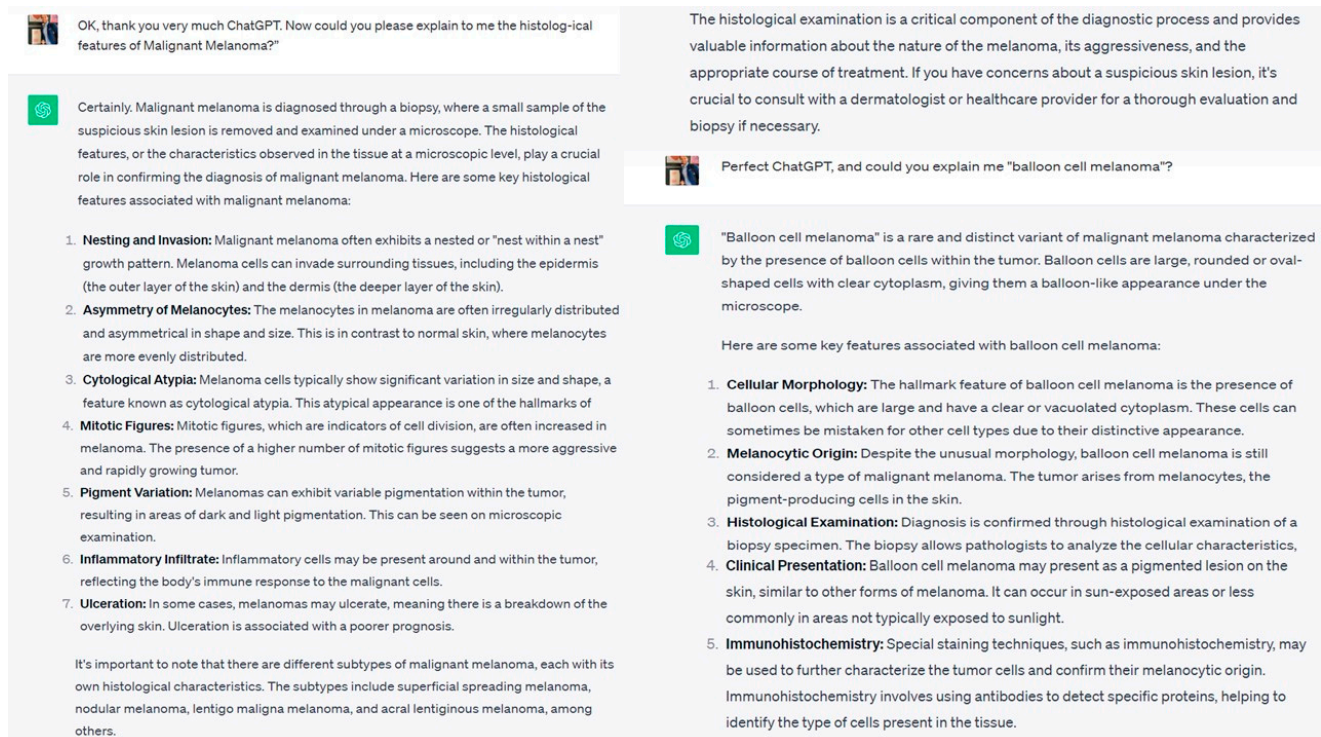


Figure 4. Other answers of ChatGPT to questions proposed by authors.



Could be possible to know how to differentiate the gleason score of the prostate adenocarcinoma?



The Gleason score is a system used to grade prostate adenocarcinoma, which is the most common type of prostate cancer. It is a key component in assessing the aggressiveness of the cancer based on its microscopic appearance. The Gleason score is determined by evaluating the two most predominant architectural patterns observed in the prostate biopsy or surgical specimen.

Here's a brief overview of how the Gleason score is interpreted:

1. **Primary Grade (1-5):** The pathologist identifies the most prevalent cancer pattern in the specimen. This is assigned a grade from 1 to 5, with 1 being the least aggressive and 5 being the most aggressive.
2. **Secondary Grade (1-5):** The second most prevalent cancer pattern is identified, and it is also assigned a grade from 1 to 5.

The two grades are then added together to obtain the Gleason score, which ranges from 2 to 10. For example, if the primary grade is 3 and the secondary grade is 4, the Gleason score would be 7.

Here's a breakdown of the Gleason scores and their clinical significance:

- **Gleason Score 6 or Less:** This is considered low-grade cancer and is often associated with a more favorable prognosis. It may be referred to as well-differentiated or low-risk prostate cancer.
- **Gleason Score 7:** This is an intermediate-grade cancer. Gleason 7 cancers are further subdivided into 3+4 (favorable intermediate risk) and 4+3 (unfavorable intermediate risk). The presence of a higher primary grade (4) is associated with a slightly worse prognosis.
- **Gleason Score 8-10:** These are high-grade cancers and are associated with a higher risk of aggressive behavior and poorer outcomes. Gleason 8 cancers have a primary and secondary grade of 4, and Gleason 9 and 10 cancers have a primary grade of 5.



Now, could be possible to know the histological criteria for the differential diagnosis between germ cell tumour and serous high grade carcinoma of the ovary?



Distinguishing between a germ cell tumor and high-grade serous carcinoma of the ovary involves a careful examination of histological features. Both types of ovarian tumors have distinct characteristics that pathologists use to make a diagnosis. Here are some histological criteria that help in the differential diagnosis:

Germ Cell Tumor:

1. **Histological Diversity:** Germ cell tumors often display diverse histological patterns, including dysgerminoma, yolk sac tumor, embryonal carcinoma, choriocarcinoma, and teratoma. The presence of these elements can aid in identifying germ cell tumors.
2. **Clear Cell Borders:** Some germ cell tumors, such as dysgerminoma, may have clear cell borders, resembling large, round cells with clear cytoplasm.
3. **Syncytiotrophoblastic Cells:** In cases of choriocarcinoma, the presence of syncytiotrophoblastic cells, which form multinucleated giant cells, is a characteristic feature.
4. **Schiller-Duval Bodies:** Yolk sac tumors may exhibit Schiller-Duval bodies, which are glomerulus-like structures with a central blood vessel surrounded by germ cells.
5. **Primitive Embryonal Components:** Embryonal carcinoma components may show primitive, undifferentiated cells.

High-Grade Serous Carcinoma:

1. **Complex Architectural Patterns:** High-grade serous carcinoma typically exhibits complex architectural patterns, including papillary structures, glandular formations, and solid areas.
2. **High Nuclear Grade:** The nuclei of serous carcinoma cells are often large, hyperchromatic, and pleomorphic, indicating a high nuclear grade.
3. **Psammoma Bodies:** Psammoma bodies, which are concentric, lamellated calcifications, may be present in high-grade serous carcinoma.
4. **Nuclear Atypia:** The tumor cells in high-grade serous carcinoma often show marked nuclear atypia, including irregularities in size and shape.
5. **Infiltrative Growth Pattern:** Serous carcinoma tends to infiltrate the ovarian stroma, and it may exhibit destructive stromal invasion.

Figure 5. Other query sessions in which the first author (G.C.) ask to ChatGPT “..how to differentiate the gleason score of the prostate adenocarcinoma” and “..to know the histological criteria for the differential diagnosis between germ cell tumour and serous high grade carcinoma of the ovary”.



Would be possible to know the immunohistochemical markers and molecular signature of the pleural mesothelioma?



Certainly, the diagnosis of pleural mesothelioma often involves a combination of histological examination, immunohistochemistry (IHC), and sometimes molecular testing. Immunohistochemical markers play a crucial role in distinguishing mesothelioma from other malignancies, particularly adenocarcinomas of the lung and other cancers involving the pleura. Additionally, molecular studies can provide further insight into the molecular signature of mesothelioma. Here are some commonly used immunohistochemical markers and molecular features associated with pleural mesothelioma:

Immunohistochemical Markers:

1. **Calretinin:** Calretinin is a calcium-binding protein that is often strongly expressed in mesothelioma and is considered one of the most sensitive markers for this diagnosis.
2. **WT-1 (Wilms Tumor 1):** WT-1 is a nuclear transcription factor that is expressed in mesothelioma and can help distinguish it from lung adenocarcinoma.
3. **Cytokeratin 5/6:** This cytokeratin is commonly positive in mesothelioma and is often used in conjunction with other markers for diagnostic confirmation.
4. **D2-40 (Podoplanin):** D2-40 is a marker for lymphatic endothelium and is often positive in mesothelioma, aiding in the differentiation from lung adenocarcinoma.
5. **Mesothelin:** Mesothelin is a glycoprotein that can be elevated in the serum of patients with mesothelioma, and its expression can be detected immunohistochemically.
6. **BAP1 (BRCA1-Associated Protein 1):** Loss of BAP1 expression is observed in a subset of mesotheliomas and may be associated with a more aggressive clinical course.

Molecular Features:

1. **BAP1 Inactivation:** BAP1 is a tumor suppressor gene, and its inactivation, often through genetic mutations or deletions, is associated with a subset of mesotheliomas. BAP1 loss is more common in epithelioid mesotheliomas.
2. **CDKN2A/ARF Deletion:** Deletion of the CDKN2A/ARF locus on chromosome 9p21 is frequently observed in mesothelioma. This deletion can be detected by fluorescence in situ hybridization (FISH) or other molecular techniques.
3. **NF2 (Neurofibromin 2) Alterations:** NF2 mutations or loss are found in a subset of mesotheliomas, particularly in the sarcomatoid subtype.
4. **Mutation in TERT Promoter:** Mutations in the promoter region of the telomerase reverse transcriptase (TERT) gene are observed in some mesotheliomas.



Now, what are the molecular features of diffuse midline glioma?



Diffuse midline gliomas (DMGs) are a group of aggressive and infiltrative brain tumors that typically arise in the midline structures of the central nervous system, such as the thalamus, brainstem, or spinal cord. These tumors are most commonly diagnosed in pediatric patients. Molecular studies have provided insights into the genetic alterations associated with diffuse midline gliomas. The majority of DMGs are now classified based on the presence of specific genetic mutations and alterations. The most common molecular features associated with diffuse midline gliomas include:

1. **H3 K27M Mutation:** A hallmark molecular feature of many diffuse midline gliomas is the presence of a specific histone mutation, known as the H3 K27M mutation. This mutation affects the H3F3A or HIST1H3B/C genes, resulting in a substitution of lysine (K) with methionine (M) at position 27 of the histone H3 protein. The H3 K27M mutation is often found in midline gliomas with a histone H3.3 mutation.
2. **Histone H3.3 Mutation:** In addition to the H3 K27M mutation, some diffuse midline gliomas have mutations affecting the H3F3A gene, which encodes histone H3.3. These mutations can occur in other positions of the H3 protein and are associated with distinct clinicopathological features.
3. **IDH Wild-Type:** Unlike other gliomas, diffuse midline gliomas typically do not harbor mutations in the isocitrate dehydrogenase (IDH) genes. As a result, they are classified as IDH wild-type.
4. **TP53 Mutation:** TP53 mutations are often observed in diffuse midline gliomas, particularly in those with H3 K27M mutations. TP53 alterations contribute to the tumor's aggressive behavior.
5. **ACVR1 Mutation:** In a subset of diffuse midline gliomas, particularly those arising in the thalamus, mutations in the ACVR1 gene have been identified. These mutations are associated with the development of a specific histological variant known as diffuse midline glioma, H3 K27M-mutant with histone-wildtype features.
6. **PDGFRA Alterations:** Some diffuse midline gliomas, particularly those in the spinal cord, may show alterations in the PDGFRA gene, which encodes the platelet-derived growth factor receptor alpha.

Figure 6. Another query session related to immunohistochemical features and molecular characteristics of pleural mesothelioma and to molecular features of pediatric diffuse midline glioma (diffuse intrinsic pontine glioma).

3. Results

A total of 103 records were initially identified in the literature search, of which 19 were duplicates. After screening for eligibility and inclusion criteria, only five publications were ultimately included (Figure 1). The majority of publications were original articles ($n = 2$), followed by a case report ($n = 1$), letter to the editor ($n = 1$) and review ($n = 1$).

Figure 1 summarizes the review process following PRISMA guidelines and Table 1 summarizes the features of the five studies included in our review.

Table 1. Features (name of author, year, number of reference, type of application of ChatGPT, strengths and weaknesses) of ChatGPT in the publication analyzed in our literature review.

Year Authors	Type of Paper	Application of ChatGPT	Strengths	Weakness
Sallam 2023 [12]	Review	Scientific research	Speeding of the review Computer code generation	Erroneous contents Hallucination phenomena
		Medical practice	Simplification of the workflow	Risk of incorrect/inaccurate information
			Improved diagnostics, cost savings, improved health literacy	Transparency and legal issues
		Health education		Limited knowledge before 2021 Risk of spreading misinformation Copyright issues Lack of originality
Sinha [14] 2023	Article	Query session of 100 questions	Reasonable level of rationality	Lack of true understanding of the underlying significance and context of the information
Sorin [15] 2023	Article	ChatGPT in a molecular tumor board	Clinical recommendations of ChatGPT in line with those of the oncology committee in 70% of the cases	High difficulty in providing empirical decisions on the therapeutic path
Naik [16] 2023	Case Report	ChatGPT in the setting of clinical management	Provide clinical and pathological information	Allucination phenomena
Yamin MA [17] 2023	Article	ChatGPT in gastrointestinal pathology	Summarize patient records	Risk of inaccurate information
			Incorporation into digital pathology Education and research	Allucination phenomena

Regarding the ‘query session’, the first question was related to the diagnostical and therapeutical approach in a clinical case of a ‘32 years-old woman with a pigmented lesion on the thigh’. ChatGPT was able to answer correctly and comprehensively, summarizing the different anamnestic stages peculiar to the case, with a satisfactory explanation of concepts such as dermoscopy, biopsy and main differential diagnosis. Furthermore, ChatGPT cautioned against using this information without proper vetting and validation by an experienced pathologist. We consider this first step very satisfactory (ChatGPT’s responses were assessed individually by three reviewers who checked each response to the standardized responses to evaluate the response’s consistency. Based on their ranking, responses were classified as either “consistent” or “inconsistent.” Throughout the evaluation process, there

was strong agreement amongst the three reviewers, guaranteeing a consistent and unbiased assessment of the responses).

Going into more detail, the first author went on to ask ChatGPT: “OK, thank you ChatGPT. Now, could you please explain to me the clinical criteria for suspecting malignant melanoma?” and ChatGPT, in a very performant and correct manner, replied: “Certainly, I can provide you with an overview of the clinical criteria for suspecting malignant melanoma, but please keep in mind that I am not a medical professional, and you should consult with a healthcare professional for personalized advice. The diagnosis and evaluation of skin lesions should be done by a qualified dermatologist or healthcare provider.

Malignant melanoma is a type of skin cancer that arises from melanocytes, the pigment-producing cells in the skin. Detecting melanoma involves assessing various clinical features, often using the ABCDE criteria:

Asymmetry: One half of the mole does not match the other half in terms of shape, size, or color.

Border irregularity: The edges of the mole are not well-defined; they may be uneven, notched, or blurred.

Color variations: Melanomas often have uneven coloring, with different shades of brown, black, or even red, white, or blue.

Diameter: Melanomas are typically larger than common moles. While size alone is not a definitive indicator, moles larger than 6 millimeters (about the size of a pencil eraser) should be examined.

Evolution or change in appearance: Any change in size, shape, color, or elevation, or the development of new symptoms such as bleeding or itching, should be carefully evaluated.

In addition to the ABCDE criteria, the “EFG” rule is sometimes used:

Evolving: The mole is changing in size, shape, or color.

Firm: The mole feels hard or lumpy.

Growing: The mole is increasing in size.

It’s important to note that while these criteria are helpful, not all melanomas exhibit all these characteristics, and some benign moles may exhibit similar features. Regular skin self-exams and professional skin checks are crucial for early detection”. (see Figure 3 for the rest of ChatGPT’s response).

Again, ChatGPT proved to be quite performant and was able to provide correct answers to the appropriate questions (Figure 3).

Finally, the first author asked ChatGPT: “OK, thank you very much ChatGPT. Now could you please explain to me the histological features of Malignant Melanoma?” and ChatGPT replied: “Certainly. Malignant melanoma is diagnosed through a biopsy, where a small sample of the suspicious skin lesion is removed and examined under a microscope. The histological features, or the characteristics observed in the tissue at a microscopic level, play a crucial role in confirming the diagnosis of malignant melanoma. Here are some key histological features associated with malignant melanoma:

Nesting and Invasion: Malignant melanoma often exhibits a nested or “nest within a nest” growth pattern. Melanoma cells can invade surrounding tissues, including the epidermis (the outer layer of the skin) and the dermis (the deeper layer of the skin).

Asymmetry of Melanocytes: The melanocytes in melanoma are often irregularly distributed and asymmetrical in shape and size. This is in contrast to normal skin, where melanocytes are more evenly distributed.

Cytological Atypia: Melanoma cells typically show significant variation in size and shape, a feature known as cytological atypia. This atypical appearance is one of the hallmarks of malignant cells.

Mitotic Figures: Mitotic figures, which are indicators of cell division, are often increased in melanoma. The presence of a higher number of mitotic figures suggests a more aggressive and rapidly growing tumor.

Pigment Variation: Melanomas can exhibit variable pigmentation within the tumor, resulting in areas of dark and light pigmentation. This can be seen on microscopic examination.

Inflammatory Infiltrate: Inflammatory cells may be present around and within the tumor, reflecting the body's immune response to the malignant cells.

Ulceration: In some cases, melanomas may ulcerate, meaning there is a breakdown of the overlying skin. Ulceration is associated with a poorer prognosis.

It's important to note that there are different subtypes of malignant melanoma, each with its own histological characteristics. The subtypes include superficial spreading melanoma, nodular melanoma, lentigo maligna melanoma, and acral lentiginous melanoma, among others.

The histological examination is a critical component of the diagnostic process and provides valuable information about the nature of the melanoma, its aggressiveness, and the appropriate course of treatment. If you have concerns about a suspicious skin lesion, it's crucial to consult with a dermatologist or healthcare provider for a thorough evaluation and biopsy if necessary". Then, G.C. ask to ChatGPT: "Perfect ChatGPT, and could you explain me "balloon cell melanoma"?"

ChatGPT

"Balloon cell melanoma" is a rare and distinct variant of malignant melanoma characterized by the presence of balloon cells within the tumor. Balloon cells are large, rounded or oval-shaped cells with clear cytoplasm, giving them a balloon-like appearance under the microscope.

Here are some key features associated with balloon cell melanoma:

Cellular Morphology: The hallmark feature of balloon cell melanoma is the presence of balloon cells, which are large and have a clear or vacuolated cytoplasm. These cells can sometimes be mistaken for other cell types due to their distinctive appearance.

Melanocytic Origin: Despite the unusual morphology, balloon cell melanoma is still considered a type of malignant melanoma. The tumor arises from melanocytes, the pigment-producing cells in the skin.

Histological Examination: Diagnosis is confirmed through histological examination of a biopsy specimen. The biopsy allows pathologists to analyze the cellular characteristics, including the presence of balloon cells, and determine the malignancy of the lesion.

Clinical Presentation: Balloon cell melanoma may present as a pigmented lesion on the skin, similar to other forms of melanoma. It can occur in sun-exposed areas or less commonly in areas not typically exposed to sunlight.

Immunohistochemistry: Special staining techniques, such as immunohistochemistry, may be used to further characterize the tumor cells and confirm their melanocytic origin. Immunohistochemistry involves using antibodies to detect specific proteins, helping to identify the type of cells present in the tissue (Figure 4).

Further, when ChatGPT was asked to provide references on balloon cell melanoma, two non-existent papers were provided (hallucination phenomenon), confirming what had already been reported by several papers previously.

4. Discussion

The application of AI to medicine has significantly co-assisted physicians' therapeutic decision-making processes, not replacing but instead complementing and enhancing the indispensable figure of the human [8]. The advent of ChatGPT 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 and scientific notions [11].

In a recent paper by Sallam [12], the various advantages and limitations of using ChatGPT in areas such as scientific research, practice and health education were analyzed. In particular, scientific research could really benefit from a useful and powerful tool such as ChatGPT by speeding up the process of literature reviews and computer code generation, allowing the human user to focus more on experimental design [13,18–21]; on the other hand, several authors [13,19,22] have highlighted issues of reliability of the data provided

by ChatGPT with the generation of erroneous and/or inaccurate content, phenomena of ‘hallucination’ (by which is meant the generation of erroneous content but which can be considered plausible from a scientific point of view [23]) and the bias of the answers provided by ChatGPT, which is a reflection of the dataset used in training [12]. Finally, it is important to consider that ChatGPT may generate nonexistent references, as pointed out by Chen T.J. [24] and Lubowitz [25].

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. [14] 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. One hundred 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 of four or five. This means that 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 diseases. Although it is possible to recognize schemes and categorize data, a true understanding of the underlying significance and context of the information is lacking. AI is unable to make logical judgments 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. [15], the authors discuss how ChatGPT 3.5 can also operate within the molecular tumor board, not only starting from histopathological/diagnostic data, but also integrating other key components such as the genetic or 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 tumor 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 that incomplete training (and the consequent risk of bias) may influence the performance, and thus the responses, of the chatbot.

In another recent paper by Naik H.R. et al. [16], 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. [26], 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 four publications on Chagas disease were selected from Clarivate’s Web of Science (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 for literature reviews, 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 [17], 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 a few days ago [13], 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 [27,28] and of damage resulting from inaccurate or outright incorrect content [29–32]. 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 [33].

Furthermore, it is important to say that there is need for a new framework on publication/authorship ethics in a new age of AI-sourced digital composition; it is always important to address the hallucination phenomena with a check of the user.

From an ethical point of view, it is very important to understand the issue of patients' private data, and if the use of medical clinical records is necessary, it will be important to find a way of protecting patient information.

Future Roadmaps

As highlighted in the work of Schukow et al. [13], it is imperative to outline future perspectives that the implementation of AI models will bring to the fore; first of all, it is important to consider how the use of AI methods applied to the writing of scientific articles will be managed, how to address the issue of consent and whether to modify the editorial lines of journals taking into account the use of chatbots. Secondly, it is very important to understand how ChatGPT can impact the possible option of specialization in pathological anatomy (increase or reduction) and how and to what extent there will be a need for a critical review of AI-generated content and whether or not the role of teacher/mentor can be delegated to ChatGPT.

Projecting to a future in which such systems may become more active, their integration with clinical, genetic, anamnestic, morphological and immunohistochemical data, which have always been key to pathologists' roles, will have to be screened by professionals with medical experience and knowledge, which are areas in which ChatGPT struggles the most.

Author Contributions: Conceptualization, G.C. (Gerardo Cazzato) and M.C.; methodology, G.C. (Gerardo Cazzato), F.A. and E.M.; software, G.C. (Gerardo Cazzato), F.A. and V.L.; validation, G.C. (Gerardo Cazzato), M.C., A.M. and G.I.; formal analysis, P.P. and A.M.; investigation, G.C. (Gerardo Cazzato); resources, G.C. (Gerardo Cazzato); data curation, G.C. (Gerardo Cazzato); writing—original draft preparation, G.C. (Gerardo Cazzato); writing—review and editing, G.C. (Gerardo Cazzato) and G.C. (Gennaro Cormio); visualization, G.I.; supervision, P.P., G.C. (Gennaro Cormio) 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.

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.; et al. Implementation of artificial intelligence in medicine: Status analysis and development suggestions. *Artif. Intell. Med.* **2020**, *102*, 101780. [\[CrossRef\]](#)
2. Haug, C.J.; Drazen, J.M. Artificial Intelligence and Machine Learning in Clinical Medicine, 2023. *N. Engl. J. Med.* **2023**, *30*, 1201–1208. [\[CrossRef\]](#) [\[PubMed\]](#)
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**, *16*, 1457–1465. [\[CrossRef\]](#) [\[PubMed\]](#)
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**, *9*, e46885. [\[CrossRef\]](#)
5. Bozic, J.; Tazl, O.A.; Wotawa, F. Chatbot Testing Using AI Planning. In Proceedings of the 2019 IEEE International Conference on Artificial Intelligence Testing (AITest), Newark, CA, USA, 4–9 April 2019; pp. 37–44. [\[CrossRef\]](#)
6. Adamopoulou, E.; Moussiades, L. An Overview of Chatbot Technology. *Artif. Intell. Appl. Innov.* **2020**, *584*, 373–383. [\[CrossRef\]](#)
7. Okonkwo, C.W.; Ade-Ibijola, A. Chatbots applications in education: A systematic review. *Comput. Educ. Artif. Intell.* **2021**, *2*, 100033. [\[CrossRef\]](#)
8. Khan, R.A.; Jawaid, M.; Khan, A.R.; Sajjad, M. ChatGPT-Reshaping medical education and clinical management. *Pak. J. Med. Sci.* **2023**, *39*, 605–607. [\[CrossRef\]](#)
9. Caldarini, G.; Jaf, S.; McGarry, K. A Literature Survey of Recent Advances in Chatbots. *Information* **2022**, *13*, 41. [\[CrossRef\]](#)

10. Panch, T.; Pearson-Stuttard, J.; Greaves, F.; Atun, R. Artificial intelligence: Opportunities and risks for public health. *Lancet Digit. Health* **2019**, *1*, e13–e14, Erratum in *Lancet Digit. Health* **2019**, *1*, e113. [[CrossRef](#)]
11. Mago, J.; Sharma, M. The Potential Usefulness of ChatGPT in Oral and Maxillofacial Radiology. *Cureus* **2023**, *15*, e42133. [[CrossRef](#)]
12. Sallam, M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare* **2023**, *11*, 887. [[CrossRef](#)] [[PubMed](#)]
13. 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**, *27*, 406. [[CrossRef](#)] [[PubMed](#)]
14. Sinha, R.K.; Deb Roy, A.; Kumar, N.; Mondal, H. Applicability of ChatGPT in Assisting to Solve Higher Order Problems in Pathology. *Cureus* **2023**, *15*, e35237. [[CrossRef](#)] [[PubMed](#)]
15. 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**, *9*, 44. [[CrossRef](#)] [[PubMed](#)]
16. 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**, *15*, e37587. [[CrossRef](#)]
17. Ma, Y. The potential application of ChatGPT in gastrointestinal pathology. *Gastroenterol. Endosc.* **2023**, *1*, 130–131. [[CrossRef](#)]
18. Nature editorial Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature* **2023**, *613*, 612. [[CrossRef](#)]
19. Moons, P.; Van Bulck, L. ChatGPT: Can artificial intelligence language models be of value for cardiovascular nurses and allied health professionals. *Eur. J. Cardiovasc. Nurs.* **2023**, *22*, e55–e59. [[CrossRef](#)]
20. Biswas, S. ChatGPT and the Future of Medical Writing. *Radiology* **2023**, *307*, e223312. [[CrossRef](#)]
21. Lund, B.; Wang, S. Chatting about ChatGPT: How may AI and GPT impact academia and libraries? *Library Hi Tech News* **2023**, *40*, 26–29. [[CrossRef](#)]
22. The Lancet Digital Health. ChatGPT: Friend or foe? *Lancet Digit. Health* **2023**, *5*, e112–e114. [[CrossRef](#)]
23. Cascella, M.; Montomoli, J.; Bellini, V.; Bignami, E. Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios. *J. Med. Syst.* **2023**, *47*, 33. [[CrossRef](#)] [[PubMed](#)]
24. Chen, T.J. ChatGPT and Other Artificial Intelligence Applications Speed up Scientific Writing. *J. Chin. Med. Assoc.* **2023**, *86*, 351–353. [[CrossRef](#)] [[PubMed](#)]
25. Lubowitz, J. ChatGPT, An Artificial Intelligence Chatbot, Is Impacting Medical Literature. *Arthroscopy* **2023**, *39*, 1121–1122. [[CrossRef](#)]
26. 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**, *51*, 1360–1361. [[CrossRef](#)]
27. Holzinger, A.; Keiblinger, K.; Holub, P.; Zatloukal, K.; Müller, H. AI for life: Trends in artificial intelligence for biotechnology. *New Biotechnol.* **2023**, *74*, 16–24. [[CrossRef](#)]
28. Jeblick, K.; Schachtner, B.; Dexl, J.; Mittermeier, A.; Stuber, A.T.; Topalis, J.; Weber, T.; Wesp, P.; Sabel, B.; Ricke, J.; et al. ChatGPT Makes Medicine Easy to Swallow: An Exploratory Case Study on Simplified Radiology Reports. *arXiv* **2022**, arXiv:2212.14882. [[CrossRef](#)]
29. Ahn, C. Exploring ChatGPT for information of cardiopulmonary resuscitation. *Resuscitation* **2023**, *185*, 109729. [[CrossRef](#)]
30. 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**, *92*, 663–664. [[CrossRef](#)]
31. Patel, S.B.; Lam, K. ChatGPT: The future of discharge summaries? *Lancet Digit. Health* **2023**, *5*, e107–e108. [[CrossRef](#)]
32. Ali, S.R.; Dobbs, T.D.; Hutchings, H.A.; Whitaker, I.S. Using ChatGPT to write patient clinic letters. *Lancet Digit. Health* **2023**, *5*, e179–e181. [[CrossRef](#)] [[PubMed](#)]
33. Goddard, J. Hallucinations in ChatGPT: A Cautionary Tale for Biomedical Researchers. *Am. J. Med.* **2023**, *136*, 1059–1060. [[CrossRef](#)] [[PubMed](#)]

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.