

Original Article

Artificial Intelligence in Medicine: A Comparative Study of ChatGPT and Google Bard in Clinical Diagnostics

Aso S. Muhialdeen¹, Shorsh A. Mohammed¹, Nahida Hama Ameen Ahmed², Shaho F. Ahmed¹, Wriya N. Hassan¹, Hoshmand R. Asaad³, Dana T. Gharib³, Huda M. Muhammad⁴, Shko H. Hassan¹, Karokh Fadhil Hama Hussein³, Hemin S. Mohammed¹, Abdulwahid M. Salih⁴, Fahmi H. Kakamad⁴, Muhammed Karim⁵, Fakher Abdullah⁵, Hemn A. Hassan⁵, Sasan M. Ahmed⁵, Suhaib H. Kakamad⁵, Marwan N. Hassan^{5*}, Shvan H. Mohammed⁵, Berun A. Abdalla¹

1. Scientific Affairs Department, Smart Health Tower, Madam Mitterrand Street, Sulaimani, Kurdistan, Iraq
2. Sulaimani Maternity Teaching Hospital, Sulaimani Kurdistan, Iraq
3. Gastroenterology and Hepatology Teaching Hospital, Sulaimani, Kurdistan, Iraq
4. College of Medicine, University of Sulaimani, Madam Mitterrand Street, Sulaimani, Kurdistan, Iraq
5. Kscien Organization for Scientific Research (Middle East office), Hamid Str, Azadi Mall, Sulaimani, Kurdistan, Iraq

* Corresponding author: marwan.nasih12@gmail.com (M.N. Hassan). Gullan 71, House number 6, Zip code: 46018, Halabja, Iraq



Keywords:

Artificial intelligence
ChatGPT
Google bard
Clinical diagnosis
Medical decision-making

Received: October 5, 2023

Revised: October 15, 2023

Accepted: October 25, 2023

First Published: November 6, 2023

Copyright: © 2023 Muhialdeen et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Citation: Muhialdeen AS, Mohammed SA, Ahmed NH, Ahmed SF, Hassan WN, Asaad HR et al. Artificial Intelligence in Medicine: A Comparative Study of ChatGPT and Google Bard in Clinical Diagnostics. Barw Medical Journal. 2024;2(1):7-13.
<https://doi.org/10.58742/pry94q89>

Abstract

Introduction

The introduction of Artificial Intelligence (AI) tools like ChatGPT and Google Bard promises transformative advances in clinical diagnostics. The aim of this study is to examine the ability of these two AI tools to diagnose various medical scenarios.

Methods

Experts from varied medical domains curated 20 case scenarios, each paired with its ideal diagnostic answer. Both AI systems, ChatGPT (updated in September 2021) and Google Bard (updated in January 2023), were tasked with diagnosing these cases. Their outcomes were recorded and subsequently assessed by human medical professionals.

Results

In the diagnostic evaluations, ChatGPT achieved an accuracy of 90%, correctly diagnosing 18 out of 20 cases, while Google Bard displayed an 80% accuracy rate, correctly answering 16 questions. Notably, both AIs faltered in specific complex scenarios. For instance, both systems misdiagnosed a labor situation, and while ChatGPT incorrectly identified a case of hypertrophic pyloric stenosis, Google Bard suggested a less suitable diagnostic procedure (pelvic ultrasound) for a 56-year-old patient.

Conclusion

This study showcases the promising capabilities of ChatGPT and Google Bard in the realm of clinical diagnostics, with both AI tools achieving commendable accuracy rates.

1. Introduction

The medical industry might see revolutionary shifts with the advent of artificial intelligence (AI) technologies, ushering in improved patient care and outcomes. OpenAI unveiled an innovative language processing system, known as Chat Generative Pre-Trained Transformer (ChatGPT), in November 2022, celebrated for its capability to generate human-like answers across various tasks [1,2]. Google's textual AI tool, named Google Bard, leverages advanced machine learning and linguistic analysis, showing prowess in addressing complex questions. Both these platforms (ChatGPT and Google Bard) have quickly gained global recognition owing to their remarkable skill in replicating human dialogue and thought processes. Their reliability and accuracy in medical contexts have also been validated [3].

A prominent advantage of ChatGPT within medical decision-making lies in its capability to meticulously analyze complex medical information. It can serve as a valuable resource for healthcare professionals, providing swift feedback from patient histories, research papers, and medical protocols [1,4]. An essential aspect of ChatGPT is its neutrality. Operating on data-driven algorithms, the platform can potentially sidestep human biases, offering arguably more consistent recommendations. Moreover, ChatGPT can play a crucial role in distinguishing diagnoses and amalgamating data from patient symptoms, historical records, and predispositions. Evaluating this information holistically can outline a spectrum of possible health issues, assisting doctors in their assessments. This can refine diagnostic accuracy, reducing errors or postponements [4,5].

The integration of both ChatGPT and Google Bard in medical decision-making processes has garnered attention across varied medical specialties. Multiple sectors have authored papers underscoring the value and potential applications of these two platforms in their respective fields [2,3].

The purpose of this study was to contrast the decision-making behaviors of AI and humans, exploring the potential pros and cons of embedding AI instruments into medical decisions on an individual basis.

2. Methods

A collection of case scenarios, comprising 20 questions along with their best answers, was curated by experts from various fields such as general surgery, internal medicine, gynecology, endocrinology, and pediatrics. These cases and responses underwent review by separate contributors. Subsequently, they were individually presented to ChatGPT (last updated in September 2021) and Google Bard (last updated in January 2023). The outcomes were tabulated and then evaluated by the human contributors.

3. Results

ChatGPT and Google Bard both demonstrated notable accuracy in diagnosing the cases, with Google Bard achieving 80%

accuracy on 16 questions and ChatGPT at 90% for 18 questions. However, both systems provided incorrect guidance for a challenging labor situation, recommending assisted vaginal delivery instead of a cesarean section. ChatGPT mistakenly identified a case of hypertrophic pyloric stenosis as gastroesophageal reflux. Google Bard, on the other hand, suggested a pelvic ultrasound instead of the more appropriate transvaginal ultrasound for a 56-year-old woman who noticed consistent dark brown spots on her underwear. Furthermore, when faced with a week-old infant showing symptoms of vomiting a greenish substance, though the actual diagnosis was malrotation, Google Bard incorrectly pinpointed it as bilious ileus (Table 1).

4. Discussion

The recent research that juxtaposed the diagnostic capabilities of AI tools, namely ChatGPT and Google Bard, with human healthcare professionals in outpatient settings, sheds light on the potential trajectory of modern medicine. The findings revealed that both ChatGPT and Google Bard achieved commendable accuracy rates of 90% and 80%, respectively, in immediate diagnoses. This highlights AI's aptitude to decode and associate complex patterns within vast medical information, suggesting imminent transformations in the medical sector [6,7].

A significant benefit of integrating AI in outpatient diagnostics is its objective nature. AI systems, like ChatGPT and Google Bard, rely on predetermined algorithms, which, in theory, could yield unbiased diagnostic outcomes. This lack of bias might lead to more consistent, and possibly more precise, results. However, it's crucial to mention that the controlled setting of the research may not mirror the intricacies of real-world clinical scenarios. Real-life variations in symptom manifestation, patient backgrounds, and intricate details could pose unique diagnostic hurdles [8,9].

A study by Gilson et al assessed ChatGPT's ability to answer questions from the United States Medical Licensing Examination (USMLE) Step 1 and Step 2. Using questions from both AMBOSS, a popular medical student question bank, and the National Board of Medical Examiners' free 120 questions, ChatGPT's performance was compared with other language models, GPT-3 and InstructGPT. ChatGPT achieved passing scores comparable to a third-year medical student on the NBME-Free-Step-1 data set and consistently provided logical justification for its answers. The results suggest ChatGPT's potential as an interactive tool for medical education [5].

In the non-peer-reviewed study, Ahmed and colleagues provided an in-depth comparison of the conversational AI models ChatGPT and Bard, developed by OpenAI and Google AI, respectively. The research explored their architectures, training techniques, performance metrics, and limitations, as well as delving into ethical concerns like biases and potential misuse. The results underscored ChatGPT's notable performance, marking it as a dominant model in the field. The paper serves as a crucial guide for those in the conversational AI domain, emphasizing the potential advancements and implications of these advanced language models [10].

Table 1. Full description of the cases and corresponding recommendations.

Case description	Chat GPT recommendation	Google bard recommendation	Physician's decision
A 26-year-old lady with a palpable mobile mass in the left lower cervical region, a small suspicious nodule in the left thyroid lobe, and a large suspicious lymph node (LN) in the left lower cervical region. The FNAC (Fine Needle Aspiration Cytology) of the cervical LN was positive for cancer, suggestive of metastatic Papillary Thyroid Carcinoma (PTC). What is the management plan?	Total thyroidectomy + LN dissection+ RAI.	Total thyroidectomy + RAI.	Total thyroidectomy + LN dissection+ RAI.
A 32-year-old lady with an incidental finding of a suspicious solitary small (0.5 cm) nodule in the right thyroid lobe, which was confirmed to be Papillary Thyroid Microcarcinoma (PTMC) on Fine Needle Aspiration Cytology (FNAC), what is the management plan?	Lobectomy+- completion.	Total thyroidectomy.	Right thyroid lobectomy + consent for completion thyroidectomy.
A 75-year-old female patient with multinodular goiter (MNG), significant retrosternal extension of the right thyroid lobe, euthyroid status, and a benign colloid nodule on Fine Needle Aspiration Cytology (FNAC) with no malignancy.	Thyroidectomy, if compression are symptoms present.	Thyroidectomy, if compression are symptoms present.	Total thyroidectomy.
In your antenatal clinic, there is a 34-year-old woman who has given birth once before (para 1) and has a BMI of 29. She is currently at 10 weeks of pregnancy with a confirmed pregnancy within the uterus, and she is taking a daily dose of 400 mcg of folic acid. During her previous pregnancy, which occurred three years ago, she experienced preeclampsia at 37 weeks of gestation. At that time, she underwent induction and delivered a healthy baby girl at 37+2 weeks, with a weight of 2990 grams. It's worth noting that she does not have any additional risk factors.			
A 28-year-old woman, pregnant for the first time, begins labor spontaneously at 40+6 weeks of pregnancy. The initial stage of labor is extended, starting at 5 centimeters of dilation and continuing for 11 hours. Following a passive second stage of labor lasting one hour, she actively pushes for two hours but becomes fatigued. Upon examination, the baby's head is positioned cephalically, with approximately 2/5 of it palpable through the abdominal wall. The cervix is fully dilated, and the baby's head is in a direct occiput posterior (OP) position with noticeable caput (swelling), significant molding, and a station of -1. Contractions are strong, occurring every 4 minutes and lasting 10 seconds each. The fetal heart rate monitoring (CTG) is normal, and the epidural anesthesia is effectively relieving pain. What is the most appropriate course of action?	Assisted vaginal delivery.	Assisted vaginal delivery.	Caesarean section.
A 56-year-old woman, who has a BMI of 38 and has been diagnosed with type 2 diabetes, reports a continuous occurrence of dark brown staining on her underwear for the last seven days. She underwent menopause at the age of 53 and has not experienced any additional bleeding or discharge since then. There is no history of vaginal or vulval injury, and her cervical smear test, conducted six months ago, showed no abnormalities. During a physical examination, her cervix appears normal, and there is no indication of external hemorrhoids. A urine dipstick test came back negative for the presence of blood. What would be the most likely subsequent course of action?	Transvaginal ultrasound.	Ultrasound of the pelvis.	Transvaginal ultrasound.
An 80-year-old male patient returns for a follow-up appointment concerning his inclusion body myositis condition. He has completed a four-month prednisolone taper without experiencing any clinical improvement. During a physical examination, his vital signs are within normal range. However, there has been no change in the weakness observed in his shoulders, forearms, hand grip, hip girdle, and knee extensors since the previous examination. His serum creatine kinase (CPK) level is 370 U/L, showing a slight decrease compared to four months ago. What would be the most suitable course of action for treatment?	Physical Therapy.	Immunosuppressive drugs.	Physical Therapy.
A 51-year-old woman is undergoing an immigration medical evaluation. She has had Rheumatoid Arthritis for 2 years and needs vaccinations. However, she lacks a vaccination record. She is currently taking oral methotrexate and folic acid, along with tofacitinib.	Measles, Mumps, and Rubella (MMR) Vaccine:	Measles, Mumps, and Rubella (MMR) Vaccine:	Measles, Mumps, and Rubella (MMR) Vaccine:
Which vaccines should she avoid due to her medical condition and medications?			

Table 1. Continued...

<p>A 44-year-old woman is attending a follow-up appointment to review the results of her bone mineral density test. She has been living with myasthenia gravis for a decade and has been on long-term glucocorticoid therapy. Her current medications include azathioprine, prednisolone, calcium, vitamin D, and pyridostigmine.</p> <p>The DXA scan shows a T-score of -2.1 at the lumbar spine and -1.9 at the femoral neck. Her fracture risk assessment tool score is 3.5% at the hip and 17% overall.</p> <p>What would be the most suitable treatment plan in this case?</p>	Biphosphanates.	Biphosphanates.	Biphosphanates.
<p>A 4-year-old boy reported blood in his urine to his mother, who then confirmed the presence of hematuria. The mother sought medical evaluation for her child. The boy denied any recent significant injuries and had been generally healthy. His medical history was unremarkable. He falls within the 56th percentile for height and the 43rd percentile for weight. Upon physical examination, the boy appeared healthy, with normal vital signs. Cardiopulmonary examination findings were within the expected range. An abdominal mass measuring 10 centimeters in size was discovered in the left upper quadrant. This mass was firm and not painful. No abnormalities were observed in the extremities. Laboratory tests, including a complete blood count (CBC) and electrolyte levels, showed results within the normal range. Urinalysis indicated the presence of 50 to 100 red blood cells per high-power field.</p> <p>What is the most probable diagnosis?</p>	Wilms Tumor (Nephroblastoma).	Wilms Tumor (Nephroblastoma).	Wilms Tumor (Nephroblastoma).
<p>A 9-month-old boy, who was previously in good health, is taken to the emergency department due to severe, on-and-off abdominal pain. These episodes of pain happen every 10 to 15 minutes. During these painful episodes, the child pulls his legs towards his abdomen. Additionally, the child is experiencing vomiting and has stools with visible blood. Upon physical examination, a tender, movable, tubular mass is detected in the middle of the abdomen. What is the most probable diagnosis?</p>	Intussusception	Intussusception	Intussusception
<p>A 15-year-old adolescent arrives at the emergency department with a temperature of 38.9°C (102°F), a solid and immobile lump in the right lower part of the abdomen, and the primary concern of abdominal discomfort. The patient has been unwell for the past two weeks but had not sought medical attention until now. What is the probable diagnosis?</p>	Appendicitis with Abscess Formation.	Appendicitis.	Appendicitis with Abscess Formation.
<p>While working in the emergency room, a mother contacts you regarding her 5-week-old infant who has been experiencing vomiting. The baby was born at full term without any complications during birth and is exclusively breastfed. The vomiting episodes began two days ago and now happen after each feeding. What might be the likely medical diagnosis?</p>	Gastroesophageal Reflux.	Hypertrophic pyloric stenosis.	Hypertrophic pyloric stenosis.
<p>A 45-year-old male has come to the medical facility with complaints of loss of appetite, feelings of nausea, vomiting, and substantial weight loss. His medical history includes hypertension, ischemic heart disease, and osteoarthritis in both knees. He has been taking antihypertensive medications (amlodipine 5mg and valsartan 160mg), aspirin 100mg, and atorvastatin 20mg. Additionally, he has been using herbal medicine to manage knee pain for the past 5 years but stopped it a month ago. His vital signs indicate low blood pressure (80/50 mmHg), an elevated heart rate (110 bpm), and a normal temperature (36.5°C).</p> <p>Next steps for investigation and treatment:</p>	<ul style="list-style-type: none"> • Blood Work (CBC, RFT, LFT, Electrolyte) • Cardiac Evaluation • Imaging • Endoscopy Treatment: • Stabilization • Cardiac Evaluation • Medication Review: Review all of his medications, including the use of herbal medicine. • Gastrointestinal Evaluation <p>Nutritional Support</p>	<ul style="list-style-type: none"> • Complete blood count (CBC) • Electrolytes • Liver function tests • Renal function tests • Urinalysis • Abdominal ultrasound • CT scan of the abdomen and pelvis • Endoscopy • Biopsy <p>Treatment according to results.</p>	<p>Investigations: CBC, LFT, RFT s. electrolytes, GUE, Glucose, ESR, CRP, TFT, B12</p> <p>Ultrasound of abdomen, S. Cortisol and short synacthen test</p> <p>Diagnosis: Adrenal insufficiency due to exogenous steroid in herbal.</p> <p>Hydrocortisone tab 10 mg twice daily started and he has responded very well to the steroid.</p>

Table 1. Continued...

A term infant who is 1 week old and previously healthy is brought in with a 1-day history of vomiting greenish fluid (bilious emesis), reduced urine output, and findings of delayed capillary refill during a physical examination. The infant has clear lung sounds on both sides, with no detectable heart murmurs. The abdomen is not swollen and doesn't appear to be tender. The most probable diagnosis in this case is:

Malrotation with Volvulus.

Bilious ileus.

Malrotation.

A young man presented with persistent right lower leg swelling over the past three years. Initially, he experienced severe pain in the lower part of his right leg after a traumatic event, leading to swelling. He sought medical care in the emergency room, where X-rays ruled out fractures, and he was provided with pain relief before being discharged. However, his symptoms persisted and evolved to include discoloration and itching of the right lower leg. Eventually, an ulcer developed above the lateral malleolus, and a skin biopsy confirmed stasis dermatitis. There are no significant findings in his medical, medication, or family history, and he is a non-smoker and non-alcoholic.

Refer the patient to a vascular specialist (DVT).

Chronic venous insufficiency (CVI).

Doppler US (Chronic venous insufficiency (CVI)).

What is the most appropriate next step in his management?

An 11-year-old boy has been referred to the Endocrinology Department due to erratic blood sugar levels. He has experienced repeated episodes of severe hypoglycemia, and on several occasions, he had to be admitted to the intensive care unit due to diabetic ketoacidosis. He has a confirmed diagnosis of type 1 diabetes mellitus and has been on a mixed insulin regimen. However, over the past year, he has been facing hypoglycemic episodes despite being on relatively low insulin doses.

Causes of Fluctuating Blood Sugar and Hypoglycemia:

Causes of Fluctuating Blood Sugar and Hypoglycemia:

Firstly, we should investigate the reasons behind the blood sugar fluctuations. This involves gathering information on insulin management, including how it's handled, stored, the method and site of injection, any changes in injection sites, and the dosage.

- Insulin Dosing Growth and Hormonal Changes
- Insulin Sensitivity
- Diet and Exercise

- Insulin resistance
- Hormonal changes
- Stress Medications
- Unhealthy lifestyle choices.

Approach to Addressing the Issue:

Approach to Addressing the Issue:

If everything appears to be in order concerning insulin, we should explore other potential causes for the fluctuations. This includes screening for additional autoimmune conditions such as Addison's disease, Celiac disease, and thyroid disorders.

What could be the reasons behind these blood sugar fluctuations, and what is the recommended course of action to address this issue?

- Comprehensive Evaluation
- Insulin Regimen Adjustment
- Blood Sugar Monitoring
- Diet and Nutrition
- Exercise Management
- Education and Support

- Check his blood sugar levels regularly.
- Carry snacks and sugary drinks with him.
- Learn the signs and symptoms of high and low blood sugar.
- Get regular medical checkups.

Regular Follow-Up.

If all these evaluations yield normal results, we can consider offering the patient an insulin infusion pump. This device allows for the adjustment of insulin doses based on continuous glucose monitoring readings.

67-year-old man arrived at his local emergency room with complaints of feeling lightheaded and experiencing occasional irregular heart rhythms. He did not report chest pain or shortness of breath. His medical history included a diagnosis of papillary thyroid carcinoma five years ago, which was initially

Discontinue Sunitinib.

Discontinue verapamil.

Discontinue Sunitinib.

Table 1. Continued...

treated with a total thyroidectomy, modified neck dissection, and adjuvant radioiodine therapy. Although his post-therapy scan showed minimal uptake in the neck, his TSH-stimulated thyroglobulin level rose to 524 ng/mL, prompting further imaging that revealed lung lesions consistent with metastatic thyroid carcinoma. Over the subsequent 12 months, these lung lesions progressively enlarged. Repeat radioiodine imaging did not show any response, leading to the initiation of oral sorafenib therapy to manage radioiodine-refractory progressive metastases. Initially, he progressed, prompting a switch in therapy to sunitinib 14 months before his recent emergency room visit. His treatment with sunitinib was marked by adverse events, including hypertension, diarrhea, and hypomagnesemia.

Two weeks before his ER visit, he sought care from his primary care physician for acute bronchitis, and antibiotic treatment with clarithromycin was started. Four days later, he returned with worsening diarrhea and was found to have high blood pressure (170/100 mm Hg) and an elevated pulse rate (96 beats per minute). To manage his symptoms, verapamil was added to his existing valsartan therapy, and the dose of diphenoxylate/atropine for diarrhea was increased.

Upon examination in the emergency room, his pulse rate was 52 beats per minute, and his blood pressure was 126/62 mm Hg. An electrocardiogram revealed sinus bradycardia with a prolonged QTc interval of 520 msec. His serum magnesium level was found to be low at 0.9 mg/dL. While preparing for an intravenous magnesium infusion, there was an acute increase in his heart rate to 140 beats per minute, and a transient episode of torsades de pointes was identified on cardiac monitoring. Consequently, an infusion of magnesium (16 meq) was initiated, leading to an improvement in the QTc interval (reduced to 470 msec) and the resolution of his tachyarrhythmia, returning to baseline sinus bradycardia.

What should be the next best course of action?

Yet, it's essential to perceive AI not as a substitute for human skill but as a supplementary asset. The irreplaceable human element in healthcare—embodying compassion, practical wisdom, and a thorough approach to patient welfare—when harmonized with AI's computational power, could elevate the diagnostic methodology. This synergistic approach can fill potential lapses in human discernment, facilitating a well-rounded patient assessment [11,12].

While the findings are encouraging, it's critical to recognize the constraints. The research, limited to a predetermined collection of cases, might overlook real-world factors like societal influences or diverse symptom manifestations. Furthermore, AI's decision-making often lacks transparency. Commonly referred to as the 'black box' dilemma, the specific logic behind an AI-generated diagnosis remains elusive, posing challenges in medical situations where comprehending the rationale is essential [9].

5. Conclusion

The impressive performance of ChatGPT and Google Bard in this research suggests potential avenues for a more detailed exploration of AI's role in psychiatric evaluations. As tech advancements fine-tune these algorithms, the distinctions between human and AI-based diagnostic choices could blur, signifying a transformative phase in psychiatric treatment. In conclusion, while the future seems bright, it's imperative to approach the evolving AI-medical nexus with a balance of eagerness, cooperation, and prudence.

Declarations

Conflicts of interest: The author(s) have no conflicts of interest to disclose.

Ethical approval: Not applicable.

Patient consent (participation and publication): Not applicable.

Funding: The present study received no financial support.

Acknowledgements: None to be declared.

Authors' contributions: ASM and AMS major contribution of the idea. SAM, NHA, SFA, WNH, HRA and SHK participated in data collection. MNH and BAA designed the study. SHM, SHK and SMA performed the data analysis. FHK and MK participated in preparing the manuscript. HAH, KFH, HSM and FA critically revised the manuscript; all authors approved the final version of the manuscript.

Data availability statement: Note applicable.

References

1. Amisha, Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine. *Journal of Family Medicine and Primary Care*. 2019;8(7):2328–31. [doi:10.4103/jfmpc.jfmpc_440_19](https://doi.org/10.4103/jfmpc.jfmpc_440_19)
2. Salih AM, Mohammed BA, Hasan KM, Fattah FH, Najmadden ZB, Kakamad FH, et al. Mitigating the Burden of meningitis outbreak; ChatGPT and Google Bard Recommendations for the general populations; general practitioners and pediatricians. *Barw Medical Journal*. 2023;1(2). [doi:10.58742/BMJ.V1I2.32](https://doi.org/10.58742/BMJ.V1I2.32)
3. AYDIN Ö. Google Bard generated literature review: metaverse. *Journal of AI*. 2023;7(1): 1-14. [doi: N/A](https://doi.org/10.12669/pjms.39.2.7653)
4. Khan RA, Jawaid M, Khan AR, Sajjad M. ChatGPT - Reshaping medical education and clinical management. *Pakistan Journal of Medical Science*. 2023;39(2):605-7. [doi:10.12669/pjms.39.2.7653](https://doi.org/10.12669/pjms.39.2.7653)
5. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How Does ChatGPT Perform on the United States Medical Licensing Examination? The Implications of Large Language Models for Medical Education and Knowledge Assessment. *JMIR Medical Education*. 2023;9:e45312. [doi:10.2196/45312](https://doi.org/10.2196/45312)
6. Srivastav S, Chandrakar R, Gupta S, Babhulkar V, Agrawal S, Jaiswal A, et al. ChatGPT in Radiology: The Advantages and Limitations of Artificial Intelligence for Medical Imaging Diagnosis. *Cureus*. 2023;15(7):e41435. [doi:10.7759/cureus.41435](https://doi.org/10.7759/cureus.41435)
7. Baumgartner C. The potential impact of ChatGPT in clinical and translational medicine. *Clinical and Translational Medicine*. 2023;13(3):e1206. [doi:10.1002/ctm2.1206](https://doi.org/10.1002/ctm2.1206)
8. Suhag A, Kidd J, McGath M, Rajesh R, Gelfinbein J, Cacace N, Monteleone B, Chavez MR. ChatGPT: a pioneering approach to complex prenatal differential diagnosis. *American Journal of Obstetrics & Gynecology MFM*. 2023;5(8). [doi:10.1016/j.ajogmf.2023.101029](https://doi.org/10.1016/j.ajogmf.2023.101029)
9. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medical*. 2019;25(1):44–56. [doi:10.1038/s41591-018-0300-7](https://doi.org/10.1038/s41591-018-0300-7)
10. Ahmed I, Kajol M, Hasan U, Datta PP, Roy A, Reza MR. ChatGPT vs. Bard: A Comparative Study [Internet]. *TechRxiv* [Preprint]. 2023. [doi:10.36227/techrxiv.23536290.v2](https://doi.org/10.36227/techrxiv.23536290.v2)
11. Jha S, Topol EJ. Adapting to Artificial Intelligence: Radiologists and Pathologists as Information Specialists. *JAMA*. 2016;316(22):2353-54. [doi:10.1001/jama.2016.17438](https://doi.org/10.1001/jama.2016.17438)
12. Castelvecchi D. Can we open the black box of AI?. *Nature News*. 2016;538(7623):20-3. [doi:10.1038/538020a](https://doi.org/10.1038/538020a)