Evaluating the Precision of ChatGPT Artificial Intelligence in Emergency Differential Diagnosis

Abdullah Saleh Altamimi, Abdullah Ibrahim Aldughaim, Shahad Abdullah Alotaibi, Jumana Abdulqader Alrehaili, Mohamad Bakir, Ahmad Muhammad Almuhainy

***Abstract*- Introduction: artificial intelligence (AI) is the study and development of intelligent machines that can carry out tasks that would typically require human intelligence. AI seeks to give machines the ability to think, problem-solve, sense their surroundings, and comprehend human speech. By enhancing and optimising processes, this technology is predicted to completely transform a number of industries. Artificial intelligence is tipped to be the next technological breakthrough that will shape our future.**

Objective: This study focused on evaluating the precision of ChatGPT artificial intelligence in emergency differential diagnosis.

Methods: This was a comparison study, conducted from August to September 2023, evaluating the ability of both the Monica ChatGPT and the emergency medicine textbooks to provide differential diagnoses for frequently occurring complaints. Twelve symptoms common to adult patients were included in the list of chief complaints. To gauge the accuracy of the ChatGPT's answers, the researcher employed ChatGPT®-4 queries.

Results: The total number of differential diagnoses captured by the two resources was 431. The ChatGPT captured a total of 272 differential diagnoses; however, 59 of these were not included in the list of the chief complaints.

Conclusion: The study concludes that AI can be AI and human diagnosis can be used concurrently in the health sector.

***Index Terms*— Artificial Intelligence, ChatGPT, Differential Diagnoses, Emergency, Evaluation, Precision.**

Abdullah Saleh Altamimi, Abdullah Ibrahim Aldughaim and Ahmad Muhammad Almuhainy are with the Emergency Department, King Fahd Medical City, Riyadh, Saudi Arabia; email: Abdullahsaleh1892@gmail.com, Abdullah4-go@hotmail.com, aalmuhainy@kfmc.med.sa. Shahad Abdullah Alotaibi is with the College of Medicine, Sulaiman Alrajhi University, Qassim, Saudi Arabia, email: Shahadabdullah096@gmail.com. Jumana Abdulqader Alrehaili is with the College of Medicine, Taibah University, Medina, Saudi Arabia, email: Jumanax3@outlook.com. Mohamad Bakir is with the College of Medicine, Alfaisal University, email: mo7ammedbakir@gmail.com. DOI: 10.52609/jmlph.v4i1.113

I. INTRODUCTION

Artificial Intelligence (AI) is a general term that encompasses the use of a computer to model intelligent behaviour with minimal human intervention [1]. AI's main objective is to make it possible for machines to carry out cognitive tasks including problem-solving, decision-making, perception, and understanding human communication. Thus, AI-based modeling is essential for creating automated, intelligent, and smart systems that are in line with modern requirements. This technology has emerged as the next significant technological advancement, influencing the future of virtually every industry by improving, expediting, and fine-tuning their processes [2].

ChatGPT, introduced in November 2022, is an AI-based large language model (LLM) that can produce responses to text input that resemble those of a human being. Developed by OpenAI (OpenAI, L.L.C., San Francisco, CA, USA), ChatGPT is based on the generative pre-trained transformer (GPT) architecture and is referred to as a ChatGPT (a program able to interpret and generate responses using a text-based interface). The ChatGPT architecture processes natural language using a neural network, producing results based on the context of received content [2,3]. The potential applications of ChatGPTs to facilitate diagnosis and clinical judgment have been discussed previously, along with their potential benefits to personalised

medicine, drug discovery, and the analysis of enormous databases [4,5]. All of these applications, however, must be carefully assessed for potential errors encountered and mentioned in the context of LLM applications [4]. In particular, Borji thoroughly outlined the risks associated with using ChatGPT, including, but not confined to, the potential to generate erroneous information, the possibility of discrimination and prejudice, a lack of openness and dependability, cybersecurity issues, moral ramifications, and social consequences [6]. Our interaction with ChatGPT systems, which are educated and backed by human experts, is what is meant by shared expertise. This relationship results in workforce evolution, which results in the development of new capabilities [4].

A study looked at the accuracy of ChatGPT's differential diagnosis lists for clinical scenarios with typical chief complaints. For ten typical major complaints, general internal medicine physicians developed clinical cases, accurate diagnoses, and five differential diagnoses. Within the 10 differential diagnosis lists, ChatGPT correctly diagnosed 28 out of 30 cases (93.3%). Within the five differential diagnosis lists, doctors' rates of correct diagnosis were still higher than those of ChatGPT (98.3% vs. 83.3%, p = 0.03). Within the ten differential diagnosis lists produced by ChatGPT, doctors made 62/88 (70.5%) consistent differential diagnoses. In summary, the total rate of correct diagnoses within ten differential diagnosis lists generated by ChatGPT-3 was higher than 90%. This suggests that well-differentiated diagnosis lists can be developed for common chief complaints, not only by specific systems developed for diagnosis, but also by general AI ChatGPTs, such as ChatGPT-3 [7].

The effectiveness of ChatGPT in dealing with standardised clinical vignettes was studied to assess its potential for continuing clinical decision assistance. The Merck Sharpe & Dohme (MSD) Clinical Manual contains 36 published clinical vignettes that were entered into ChatGPT by the authors to examine the accuracy of differential diagnoses, diagnostic tests, final diagnoses, and management based on patient age, gender, and case acuity. The hypothetical patients portrayed in the clinical vignettes had an array of emergency severity indices (ESIs) based on the first clinical presentation, as well as a range of ages and gender identities. Throughout the 36 clinical vignettes, ChatGPT attained 71.7% overall accuracy (95% CI, 69.3% to 74.1%). It performed worse on questions involving differential diagnosis and clinical care than it did when responding to general medical knowledge questions. The authors concluded that ChatGPT performs impressively accurately when making clinical decisions, with special strengths emerging when it gets access to more clinical data [8].

Another study looked at ChatGPT's replicability and accuracy when answering questions about cirrhosis and hepatocellular carcinoma (HCC) management and emotional support. Two transplant hepatologists independently evaluated ChatGPT's solutions to 164 frequently asked questions, with a third reviewer settling any disagreements. Even while ChatGPT repeatedly recited vast amounts of information about cirrhosis and HCC, just a small portion of the accurate answers were deemed to be thorough. ChatGPT performed better in the fields of basic knowledge, lifestyle, and therapy than in diagnostic and preventative medicine. Moreover, it did not know as much about regionally specific recommendations, such as HCC screening criteria, as doctors and trainees did [9].

This study investigates the critical realm of emergency medicine, aiming to assess the precision of ChatGPT artificial intelligence in the context of differential diagnosis, with a focus on its potential to enhance diagnostic accuracy and improve patient outcomes. It also bridges a notable gap in existing literature, as prior studies have often overlooked a comprehensive evaluation of ChatGPT AI specifically tailored for differential diagnosis in emergency medicine. While advancements in AI have been explored across various medical fields, the unique challenges and exigencies of emergency scenarios necessitate a focused investigation.

II. METHODS

Study Design:

We conducted a comparison between the Monica ChatGPT and an emergency medicine textbook. For baseline information, we chose

Rosen's Emergency Medicine: Concepts and Clinical Practice, 10th Edition (2022) [10], and also used UpToDate® for baseline information, which can be accessed at https://www.wolterskluwer.com/en/solutions/uptodate. We examined the diagnostic precision of the lists of potential diagnoses created by the Monica ChatGPT for common emergency complaints, using ChatGPT®-4.

Chief complaints:

We choose 12 differential diagnoses: syncope, weakness, confusion, headache, red painful eye, diplopia, haemoptysis, chest pain, back pain, abdominal pain, constipation, and dyspnoea. These were chosen based on their being common presentations for adult patients in the emergency department.

Timeframe:

The study was conducted from August to September, 2023.

Differential Diagnosis Lists:

This study used the standard question: "What is the differential diagnosis of…?" We maintained a standardised statement and did not add any further information.

Measurements and Definitions:

We calculated the total number of differential diagnoses generated by the ChatGPT and compared it to the total number found in the textbooks. We also explored the acuity of those differentials that were captured and those that were omitted. We used only one complaint per question.

III. RESULTS

The combined number of unique differential diagnoses identified by both resources was 431. ChatGPT identified 272 of these, as detailed in Table 1. However, we observed that 59 of the identified differentials were exclusive to ChatGPT. Table 2 presents a thorough overview of the differential diagnoses identified by ChatGPT. Each row corresponds to a specific chief complaint, showcasing the wide range of conditions recognized by ChatGPT across diverse emergency medicine scenarios.

IV. DISCUSSION

This study focused on evaluating the accuracy of ChatGPT's differential diagnoses, in an emergency medicine setting, of common chief complaints as standardised in the emergency medicine textbooks referred to by the researcher. Our analysis reveals that ChatGPT successfully identifies 63.1% of the differential diagnoses. However, there are instances where ChatGPT overlooks several differential diagnoses for specific chief complaints, highlighting important limitations of this technology. We also observed that it identified an additional 59 differential diagnoses. This evaluation not only underscores areas requiring enhancement and refinement in the AI system's deployment in emergency medicine but also illustrates its ability to cover a broad range of potential diagnoses in emergency situations.

Our findings contradicted the results of Baker et al., who identified in their study that AI systems are comparable to medical professionals in terms of clinical accuracy and safety when delivering diagnostic and triage information to patients. They noted the need to start building confidence in these systems by directly comparing the performance of AI-powered systems with that of human doctors, who do not always agree on the cause of a patient's symptoms or the best course of action for triage [11]. In addition, Razzaki et al noted that the Babylon AI-powered Triage and Diagnostic System could match human physicians' precision and recall in precisely identifying the condition represented by a clinical vignette. Moreover, they found that the AI system's recommended triage was, on average, safer than that of human doctors, with only a slight decrease in appropriateness, when compared with the acceptable triage ranges provided by unbiased expert judges [12]. Another finding, by Zeltzer et al, shows that, in the context of diagnoses, there is generally strong agreement between AI and providers. The results of their study show how AI has the potential eventually to enhance patient triage and primary care disease diagnosis [13]. More so, a study by Chenais et al reveals that AI is receiving increasing attention for its potential healthcare benefits, especially in emergency medicine where

E-mail: abdullahsaleh1892@gmail.com

Table 1. Total number of differential diagnoses (DDx) captured by the AI ChatGPT

Cardinal	Total DDx in Rosen's and Up-to-date	Total DDx from ChatGPT	Total DDx Missed By ChatGPT	Total DDx Added by ChatGPT
Syncope	25	13	12	0
Weakness	43	27	24	11
Confusion	48	26	22	8
Headache	30	15	15	4
Red painful eye	25	9	16	2
Chest pain	36	27	11	5
Dyspnoea	54	36	23	10
Abdominal pain	42	32	16	7
Constipation	30	28	13	6
Back pain	42	24	24	2
Haemoptysis	24	20	7	4
Diplopia	32	15	17	0
Total	431	272	200	59

Table 2. Differential diagnoses of chief complaints detected and missed by the AI ChatGPT

Chief complaint	What was detected	What was missed
Syncope	Vasovagal	Carotid sinus hypersensitivity
	Orthostatic	Subarachnoid haemorrhage
	Medications	Mechanical fall
	Arrhythmia	Concussion
	Ischaemia	Intoxication
	Bleeding	Cataplexy
	Pulmonary embolism	Drop attacks
	Seizure	Hypertrophic obstructive
	Hypoglycaemia	cardiomyopathy
	Hypoxia	Cardiac mass
	Vertebrobasilar TIA	Tamponade
	Psychogenic pseudo-syncope	Prosthetic valve dysfunction
	Valvular heart disease	LVAD dysfunction
Weakness	Stroke	Hemiplegic migraine
	Diabetes mellitus	Todd's paralysis
	Myasthenia gravis	Hypovolaemia
	Guillain Barre Syndrome	Pre syncope
	Hypoglycaemia	Polymyalgia rheumatica
	Myositis	Systemic lupus erythematosus
	Hypokalaemia	Rheumatoid arthritis

E-mail: abdullahsaleh1892@gmail.com

	Hypercalcaemia Hypocalcaemia Hypomagnesaemia Hypophosphatemia Sepsis Acute coronary syndrome Multiple sclerosis Medications and drug abuse Rhabdomyolysis Anaemia Addison's Hypothyroidism	Temporal arteritis Brain abscess Brain tumour External compression (entrapment syndrome and compressive plexopathy) Tick paralysis Intracranial haemorrhage Subarachnoid haemorrhage Spinal cord pathology (inflammation or compression) Paraneoplastic syndromes Connective tissue disorder Vitamin deficiency Trauma Botulism Organophosphates Alcohol myopathy Thyrotoxicosis Carbon monoxide poisoning
Confusion	Parkinson's Dementia Acute coronary syndrome Arrhythmia Pulmonary embolism Brain tumour Autoimmune like SLE Multiple sclerosis	Opioid side effects/overdose Antipsychotics Sedatives Lithium Toxic alcohol Plants, Jimsonweed Parathyroid disorder Pituitary disorder Pancreas pathology Porphyria Wilson's disease Wernicke encephalopathy Vitamin B deficiency Niacin deficiency Folate deficiency Head injury Hypertensive encephalopathy Thrombocytosis Hypereosinophilia Leukemic blast cell crisis Polycythaemia Burns Electrocution Hyperthermia Hypothermia

		Trauma: with systemic inflammatory response syndrome
Headache	Stroke Vertebral artery dissection Medication side effect Chiari malformation	Post lumbar puncture Sinusitis Temporomandibular joint (TMJ) disorders Febrile headache Central Nervous System Haematoma (Epidural Haematoma and Subdural Haematoma) Brain abscess Spontaneous intracranial hypotension Idiopathic intracranial hypertension Colloid cyst Pre-eclampsia Shunt failure Traction headache Mountain sickness Anoxic headache (hypoxia)
Red painful eye	Corneal ulcer Herpes zoster ophthalmicus	Caustic injury Orbital compartment syndrome Hyphema Subconjunctival haemorrhage Corneal perforation Scleral penetration Inflamed pinguecula/pterygium Hypopyon Iritis Stye (hordeolum) Chalazion Blepharitis Contact lens overwear Dry eye syndrome Episcleritis
Diplopia	Trauma Infection/abscess Craniofacial masses Thyroid eye disease Multiple sclerosis Idiopathic intracranial hypertension Tumour Stroke Ophthalmoplegic migraine Myasthenia gravis	Wegener granulomatosis Giant cell arteritis Systemic lupus erythematosus Dermatomyositis Sarcoidosis Rheumatoid arthritis Idiopathic orbital inflammatory syndrome (orbital pseudotumor) Hypertensive vasculopathy Diabetic vasculopathy

E-mail: abdullahsaleh1892@gmail.com

	Third nerve palsy Fourth nerve palsy Sixth nerve palsy	Cavernous sinus infection, mass, vasculitis or thrombosis Orbital apex syndrome Haemorrhage Basilar artery thrombosis Vertebral artery dissection Miller Fisher or Guillain Barré syndrome Wernicke encephalopathy Botulism Tolosa-Hunt syndrome Internuclear ophthalmoplegia
Haemoptysis	Pulmonary artery aneurysm Cystic fibrosis Pseudohaemoptysis	Fungal infection Aortic aneurysm Pulmonary hypertension Thrombocytopenia Endocarditis Cocaine use Systemic lupus erythematosus
Back pain	Paraspinal muscle injury Functional back pain	Bacterial endocarditis Pulmonary embolism Pneumonia Pleural effusion Myocardial infarction Oesophageal disease Cholelithiasis (biliary colic) Cholecystitis cholangitis Perinephric abscess Ovarian torsion or tumour Pregnancy Prostatitis Acute ligamentous injury Osteoporosis Osteoid osteoma Herpes zoster Retroperitoneal haemorrhage Psoas abscess Cauda equina syndrome Transverse myelitis Isolated sciatica Tethered cord Syringomyelia Vasoocclusive pain Viral myalgia
Chest pain	Myocarditis Pericarditis	Valvular heart disease Aortic stenosis

	Arrhythmias Bronchitis Gastritis	Mitral valve prolapse Cardiac tamponade Hypertrophic cardiomyopathy Oesophageal rupture (Boerhaave syndrome) Oesophageal-tear (Mallory Weiss) Nonspecific chest wall pain Spinal root compression Postherpetic neuralgia Heart failure
Abdominal pain	Porphyria Fitz-Hugh-Curtis syndrome Rib fracture or contusion Testicular torsion Prostate-related pathology Renal infraction Musculoskeletal pain	Dissecting or ruptured aneurysm Inflammatory bowel disease Biliary colic Gastroesophageal reflux disease Hepatomegaly due to Congestive Heart Failure Myocardial ischaemia Pericarditis Myocarditis Pleural effusion Meckel's diverticulum Cecal diverticulitis Aortic aneurysm Endometriosis Psoas abscess Tubo ovarian abscess Mittelschmerz
Constipation	Faecal impaction Hernias Chagas disease Uraemia Depression Conversion disorder	Imperforate anus Anorectal atresia Aganglionosis Cerebrovascular accident Hypokalaemia Hypomagnesaemia Amyloidosis Rectocele Rectal prolapse Rectal abscess Abuse (psychological, physical, sexual) Affective disorders Postoperative pain
Dyspnoea	Arrhythmias Acute respiratory distress syndrome Bronchitis Croup	Noncardiogenic pulmonary oedema Cor pulmonale Ventilator failure Pericarditis Hypotension

E-mail: abdullahsaleh1892@gmail.com

	Vocal cord dysfunction	Sepsis from ruptured viscus
	Methemoglobinemia	Bowel obstruction
	Hypokalaemia	Toxic ingestion
	Hypocalcaemia	Renal failure
	Costochondritis	Haemothorax
	High altitude	Flail chest
		Acute chest syndrome
		Cerebral vascular accident, intracranial insult
		Multiple sclerosis
		Organophosphate poisoning
		Tick paralysis
		Neoplasm
		Cardiomyopathy
		Somatisation disorder
		Fever
		Thyroid disease
		Polymyositis
		Porphyria

many applications are currently being used. They noted that there are few studies on the various model types and validation processes, and they found no evidence for symptom checkers with decreasing performance over time. They concluded that, all things considered in the field of AI-based emergency medicine applications, there are insufficient rigorous, independent derivation, validation, or impact evaluations [14]. The AI-powered diagnostic tools, such as Babylon AI, are capable of diagnosing medical conditions with an accuracy and recall comparable to that of human doctors. Furthermore, despite its lower level of appropriateness, the assessment of the recommended AI system was found to be safer than that of human doctors. This shows that AI may improve primary care illness diagnosis and patient triage.

In line with the findings obtained in this study, a study by Rojas-Carabali et al found that ophthalmologists outperformed ChatGPT (60%) in terms of probable diagnosis, while in terms full and partial accuracy of the diagnoses, ophthalmologists achieved 76–100% success and ChatGPT achieved 72%. ChatGPT and the ophthalmologists agreed on the diagnosis in 48% of cases, and agreed on the treatment plan in 91.6% of cases. The study suggests that AI ChatGPTs can be used to diagnose and treat uveitis, and that AI can help in significantly reducing diagnostic errors [15]. On the same note, a study by Delshad et al supported this finding in its claim that the use of AI-based applications to improve the appropriateness and safety of medical triage has the potential to improve patient outcomes and experiences, as well as efficiency of healthcare delivery. They added that AI-powered applications may also be able to help with triage in more rural or underserved areas where access to traditional triage nursing services may be limited [16]. Similar to the findings revealed by the present study, a study by Rojas-Carabali et al showed that uveitis experts correctly diagnosed all cases (100%), in contrast to ChatGPT's diagnostic success rate of 66% and Glass 1.0's 33%. The study noted that the majority of participants were enthusiastic or optimistic about using AI in ophthalmology practice. It also revealed that specialists in the older age bracket and with a higher level of education had a greater proclivity to use AI-based tools. Finally, it demonstrated that ChatGPT has promising diagnostic capabilities in

uveitis cases, and that ophthalmologists had expressed interest in incorporating AI into clinical practice [17].

Based on these findings, the use of AI may be advantageous in certain diagnoses, although it is not superior to human diagnosis in most of the cases. Thus, AI can be used to aid diagnosis in areas where it appears to have an advantage over human diagnosis, but should not be used as a substitute for human diagnosis.

This research has certain limitations. It utilised a differential diagnosis for emergency cases, developed by a team of medical professionals, thus restricting the generalisability of the findings to a broader context. Furthermore, the interaction of real patients with the AI-based application may not result in the same triage decision for a given presentation as that of a specialist in the specific research field.

Various ethical considerations must also be addressed. Firstly, artificial intelligence (AI) is continuously advancing, and its capacity to offer precise differential diagnoses may evolve beyond the scope of our current study. Hence, it is prudent to regard AI as a supportive tool in diagnostic procedures, rather than relying on it exclusively. Additionally, while our research focused on a single chief complaint, incorporating multiple complaints can enhance the accuracy and credibility of the differential diagnosis. While AI aids in memorisation and suggesting differentials, it cannot supplant the critical thinking of a physician. It is crucial to recognise that AI is not authorised to provide patient treatment and should only serve as a complement to our clinical decision-making process.

V. CONCLUSION

AI-powered diagnostic tools have the potential to significantly improve patient triage and primary care illness diagnosis. Although AI's diagnostic capabilities may not always be better than those of a human, it can nevertheless be useful in certain situations. It is notable that the assessment of the recommended AI system was found to be safer than that of human physicians, implying that AI may occasionally improve patient safety. However, it is critical to recognise that AI should not be viewed as a replacement for human diagnosis, but rather as a useful tool that can supplement and enhance the skills and expertise of healthcare professionals. Further research and development is required to fully realise the potential of AI systems for healthcare and to ensure their safe and effective integration into clinical practice.

VI. REFERENCES

1. Hamet P, Tremblay J. Artificial intelligence in medicine. Metabolism. 2017 Apr;69S:S36-S40. doi: 10.1016/j.metabol.2017.01.011. Epub 2017 Jan 11. PMID: 28126242.

2. Sarker IH. AI-Based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems. SN Comput Sci. 2022 Mar 10;3(2):158.

3. Sallam M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. Healthcare. 2023 Mar 19;11(6):887.

4. Johnson KB, Wei W, Weeraratne D, Frisse ME, Misulis K, Rhee K, et al. Precision Medicine, AI, and the Future of Personalized Health Care. Clin Transl Sci. 2021 Jan 12;14(1):86–93.

5. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med. 2022 Jan 20;28(1):31–8.

6. Ali Borji. Cornell University. 2023. A Categorical Archive of ChatGPT Failures.

7. Hirosawa T, Harada Y, Yokose M, Sakamoto T, Kawamura R, Shimizu T. Diagnostic Accuracy of Differential-Diagnosis Lists Generated by Generative Pretrained Transformer 3 ChatGPT for Clinical Vignettes with Common Chief Complaints: A Pilot Study. Int J Environ Res Public Health. 2023 Feb 15;20(4):3378.

8. Rao A, Pang M, Kim J, Kamineni M, Lie W, Prasad AK, et al. Assessing the Utility of ChatGPT Throughout the Entire Clinical Workflow: Development and Usability Study. J Med Internet Res. 2023 Aug 22;25:e48659.

9. Yeo YH, Samaan JS, Ng WH, Ting PS, Trivedi H, Vipani A, et al. Assessing the performance of ChatGPT in answering questions

regarding cirrhosis and hepatocellular carcinoma. Clin Mol Hepatol. 2023 Jul 1;29(3):721–32.

10. Walls RM, Hockberger RS, Gausche-Hill M, Rosen P, editors. Rosen's emergency medicine: concepts and clinical practice. Philadelphia, PA: Elsevier; 2023.

11. Baker A, Perov Y, Middleton K, Baxter J, Mullarkey D, Sangar D, Butt M, DoRosario A, Johri S. A Comparison of Artificial Intelligence and Human Doctors for the Purpose of Triage and Diagnosis. Front Artif Intell. 2020 Nov 30;3:543405. doi: 10.3389/frai.2020.543405. PMID: 33733203; PMCID: PMC7861270.

12. Razzaki S, Baker A, Perov Y, Middleton K, Baxter J, Mullarkey D, et al. A comparative study of artificial intelligence and human doctors for the purpose of triage and diagnosis. 2018 [cited 2024 Feb 28]; Available from: https://arxiv.org/abs/1806.10698

13. Zeltzer D, Herzog L, Pickman Y, Steuerman Y, Ber RI, Kugler Z, et al. Diagnostic Accuracy of Artificial Intelligence in Virtual Primary Care. Mayo Clinic Proceedings: Digital Health. 2023 Dec;1(4):480–9.

14. Chenais G, Lagarde E, Gil-Jardiné C. Artificial Intelligence in Emergency Medicine: Viewpoint of Current Applications and Foreseeable Opportunities and Challenges. J Med Internet Res. 2023 May 23;25:e40031.

15. Rojas-Carabali W, Cifuentes-González C, Wei X, Putera I, Sen A, Thng ZX, et al. Evaluating the Diagnostic Accuracy and Management Recommendations of ChatGPT in Uveitis. Ocul Immunol Inflamm. 2023 Sep 18;1–6.

16. Delshad S, Dontaraju VS, Chengat V. Artificial Intelligence-Based Application Provides Accurate Medical Triage Advice When Compared to Consensus Decisions of Healthcare Providers. Cureus. 2021 Aug 6;13(8):e16956. doi: 10.7759/cureus.16956. PMID: 34405077; PMCID: PMC8352839.

17. Rojas-Carabali W, Sen A, Agarwal A, Tan G, Cheung CY, Rousselot A, et al. ChatGPTs Vs. Human Experts: Evaluating Diagnostic Performance of ChatGPTs in Uveitis and the Perspectives on AI Adoption in Ophthalmology. Ocul Immunol Inflamm. 2023 Oct 13;1–8.