



Breaking barriers: can ChatGPT compete with a shoulder and elbow specialist in diagnosis and management?



Mohammad Daher, BS^{a,*}, Jonathan Koa, BS^b, Peter Boufadel, BS^b, Jaspal Singh, BSc^b, Mohamad Y. Fares, MD^b, Joseph A. Abboud, MD^b

^aHotel Dieu de France, Beirut, Lebanon

^bRothman Institute/Thomas Jefferson Medical Center, Philadelphia, PA, USA

ARTICLE INFO

Keywords:

ChatGPT
Shoulder and elbow surgery
Artificial intelligence
Diagnosis and management
Empathy
Orthopedics

Level of evidence: Basic Science Study;
Validation of Computer Model

Background: ChatGPT is an artificial intelligence (AI) language processing model that uses deep learning to generate human-like responses to natural language inputs. Its potential use in health care has raised questions and several studies have assessed its effectiveness in writing articles, clinical reasoning, and solving complex questions. This study aims to investigate ChatGPT's capabilities and implications in diagnosing and managing patients with new shoulder and elbow complaints in a private clinical setting to provide insights into its potential use as a diagnostic tool for patients and a first consultation resource for primary physicians.

Methods: In a private clinical setting, patients were assessed by ChatGPT after being seen by a shoulder and elbow specialist for shoulder and elbow symptoms. To be assessed by the AI model, a research fellow filled out a standardized form (including age, gender, major comorbidities, symptoms and the localization, natural history, and duration, any associated symptoms or movement deficit, aggravating/relieving factors, and x-ray/imaging report if present). This form was submitted through the ChatGPT portal and the AI model was asked for a diagnosis and best management modality.

Results: A total of 29 patients with 15 males and 14 females, were included in this study. The AI model was able to correctly choose the diagnosis and management in 93% (27/29) and 83% (24/29) of the patients, respectively. Furthermore, of the remaining 24 patients that were managed correctly, ChatGPT did not specify the appropriate management in 6 patients and chose only one management in 5 patients, where both were applicable and dependent on the patient's choice. Therefore, 55% of ChatGPT's management was poor.

Conclusion: ChatGPT made a worthy opponent; however, it will not be able to replace in its current form a shoulder and elbow specialist in diagnosing and treating patients for many reasons such as misdiagnosis, poor management, lack of empathy and interactions with patients, its dependence on magnetic resonance imaging reports, and its lack of new knowledge.

© 2023 The Authors. Published by Elsevier Inc. on behalf of American Shoulder and Elbow Surgeons. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

ChatGPT is an artificial intelligence (AI) language processing model that employs deep learning algorithms trained on vast amounts of data to generate human-like responses to natural language inputs.¹⁹ Its capacity to process and generate natural language makes it a powerful tool for various applications, including answering questions, summarizing text, and creative writing. Amassing over 100 million users only two months after its release,²⁴ ChatGPT has gained widespread attraction and raised questions regarding its potential implementation in health care,

clinical practice, and research.^{9,23} Multiple studies are actively assessing the use of ChatGPT in writing articles, clinical reasoning, and solving complex questions,^{4,7,14,16,26} achieving impressive feats in the process, such as passing the United States Medical Licensing Examination Step 1 and Step 2 Clinical Knowledge exams.⁸

AI such as augmented reality is already finding its way into shoulder surgery.⁵ In fact, the integration of AI and automation tools such as ChatGPT in clinical practice has the potential to empower healthcare professionals and assist in the prediction, diagnosis, and treatment of a wide range of conditions. While AI has shown effectiveness in writing articles and solving clinical cases, its direct utilization in the clinical setting has yet to be explored. This study aims to investigate the capabilities and implications of ChatGPT in the diagnosis and management of patients with new

Institutional review board approval was not required for this basic science study.

*Corresponding author: Mohammad Daher, BS, Orthopedics Department, Hotel Dieu de France, Alfred Naccache Street, Beirut, Lebanon.

E-mail address: mohdaher1@hotmail.com (M. Daher).

<https://doi.org/10.1016/j.jseint.2023.07.018>

2666-6383/© 2023 The Authors. Published by Elsevier Inc. on behalf of American Shoulder and Elbow Surgeons. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

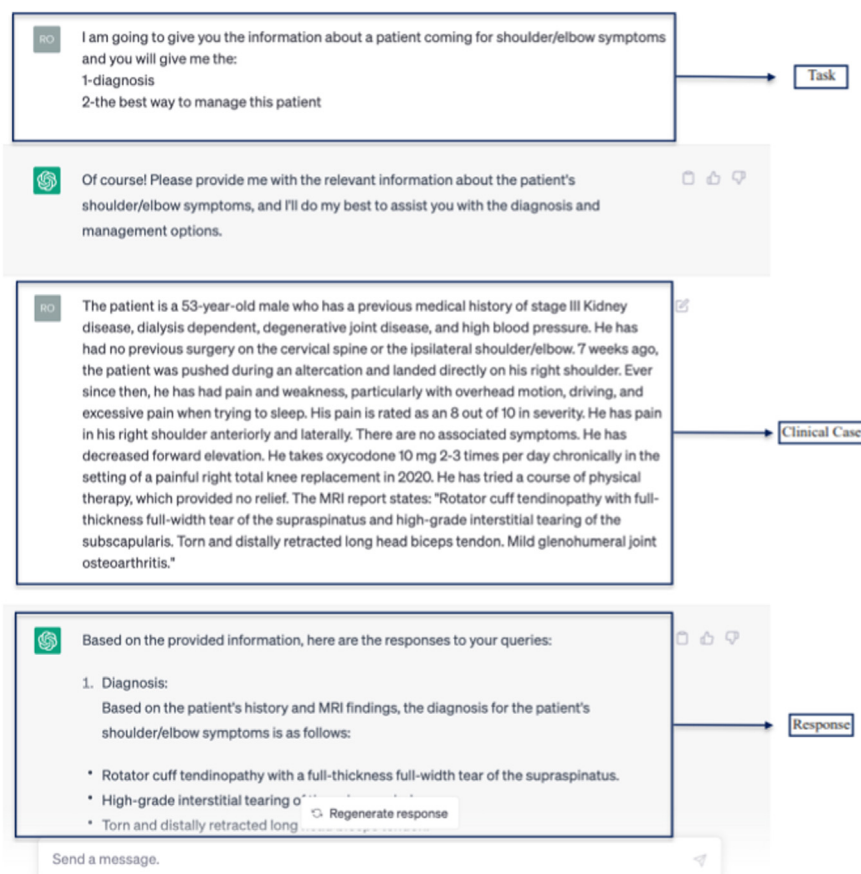


Figure 1 Example of the diagnosis and management process of ChatGPT.

shoulder and elbow complaints in a private clinical setting. We also aim to provide insights into its potential use as a diagnostic tool for patients and as a first consultation resource for primary physicians. We hypothesize that ChatGPT will perform inferiorly in comparison to a shoulder and elbow specialist in the diagnosis and management of shoulder and elbow complaints.

Material and methods

This is a prospective study of a series of 29 patients evaluated by both a shoulder and elbow specialist and ChatGPT. This study was exempt from approval by the International Review Board. First, every patient was assessed in a private clinical setting by the specialist. After that, one research fellow filled out a standardized form about the patient including age, gender, major comorbidities, symptoms and the localization, natural history, and duration, any associated symptoms or movement deficit, aggravating/relieving factors, and x-ray/imaging report if present (Fig. 1). This form was then submitted through the ChatGPT portal (chat.openai.com) and the AI model was asked for a diagnosis and a suggestion about the best management modality.

Assessment of ChatGPT's accuracy in diagnosing and management was done using descriptive statistics. The analysis included the gender ratio, the different diagnoses, rate of misdiagnosis, mismanagements, and other types of incomplete managements. The ChatGPT response was then compared to that of a fellowship-trained shoulder and elbow specialist with over 20 years of experience and was evaluated for achieving the right diagnosis, giving the right management, the comprehensiveness of the management

plan, and whether all available treatment options were included when applicable.

Results

Thirty patients were included in this study with the different diagnoses and managements by both the shoulder and elbow specialist and ChatGPT summarized in Table I.

Demographics

A total of 30 patients with 15 males and 14 females, were included in this study (Table II). These patients suffered from various shoulder pathologies with 41% (12/29) of them having rotator cuff tendinopathies, 28% (8/29) glenohumeral arthritis, 10.3% (3/29) proximal humeral fractures, 6.9% (2/29) adhesive capsulitis, 6.9% (2/29) acromioclavicular joint trauma/osteoarthritis, and 6.9% (2/30) with mild inflammation in an intact shoulder (Table II).

Misdiagnosis and mismanagement

The AI model was able to correctly choose the diagnosis and management in 93% (27/29) and 83% (24/29) of the patients, respectively (Table II). The patients of whom ChatGPT missed the diagnosis had mild inflammation in an intact shoulder. Furthermore, ChatGPT mismanaged 3 patients presenting with rotator cuff tendinopathies, 1 patient presenting with glenohumeral osteoarthritis, and 1 patient with a proximal humerus trauma making it unable to take care of 25% of rotator cuff injuries, 12.5% of

Table 1

Summary of the different diagnoses and managements by the specialist and ChatGPT.

Patient #	ChatGPT diagnosis	ChatGPT plan (conservative, surgical, or both)	ChatGPT detailed plan	Specialist diagnosis	Specialist plan (conservative surgical, or either)	Specialist detailed plan
1	<ul style="list-style-type: none"> Rotator cuff tendinopathy: a low-grade partial-thickness bursal surface insertional tear of the anterior supraspinatus tendon. Mild tendinosis of the supraspinatus and infraspinatus tendons. 	Conservative	<ul style="list-style-type: none"> Physiotherapy NSAIDs Activity modification Rest 	<ul style="list-style-type: none"> Partial thickness interstitial supraspinatus tear Possible small full-thickness anterior supraspinatus tear 	Conservative	<ul style="list-style-type: none"> Physical therapy (due to preserved strength, minimal size of tear, and encouraging physical exam)
2	<ul style="list-style-type: none"> Mild to moderate acromioclavicular joint osteoarthritis. Glenohumeral arthritis Inflammation of the rotator interval A low-grade tear of the cranial fibers of the subscapularis and subluxation of the biceps tendon into the torn subscapularis fibers. 	Conservative	<ul style="list-style-type: none"> Physical therapy NSAIDs 	<ul style="list-style-type: none"> Early glenohumeral arthritis Some inflammation of the rotator interval Some slight upper border irregularity of the subscapularis and some early subluxation of the biceps. 	Conservative	<ul style="list-style-type: none"> Physical therapy Consideration of a one-time cortisone injection (magnetic resonance imaging shows modest abnormalities and the physical exam was fairly well)
3	<ul style="list-style-type: none"> Severe end-stage glenohumeral arthritis Subscapularis tendinopathy Partial tear of the intra-articular head of the biceps tendon Mild acromioclavicular osteoarthritis 	Surgical	<ul style="list-style-type: none"> Anatomic total shoulder arthroplasty 	<ul style="list-style-type: none"> Severe glenohumeral arthritis Extensive medial erosion and bone loss 	Surgical	<ul style="list-style-type: none"> Reverse shoulder arthroplasty with a bone reconstruction system
4	<ul style="list-style-type: none"> Comminuted proximal humeral fracture 	Surgical	<ul style="list-style-type: none"> No specific option given. Possible options: open reduction internal fixation, hemiarthroplasty, and anatomic total shoulder arthroplasty 	<ul style="list-style-type: none"> Comminuted 2 part right proximal humerus fracture with varus angulation 	Surgical	<ul style="list-style-type: none"> Open reduction internal fixation
5	<ul style="list-style-type: none"> Glenohumeral arthritis 	Conservative	<ul style="list-style-type: none"> Physical therapy Pain medication Consideration of a corticosteroid injection 	<ul style="list-style-type: none"> Rotator cuff irritation Mild to moderate glenohumeral arthritis 	Conservative	<ul style="list-style-type: none"> Physical therapy
6	<ul style="list-style-type: none"> A high-grade or full-thickness tear involving the supraspinatus tendon Associated tendinosis of the rotator cuff and long head of the biceps tendon. A focal tear at the chondrolabral junction of the posterior labrum Possible capsular injury. 	Surgical	<ul style="list-style-type: none"> Arthroscopic rotator cuff repair 	<ul style="list-style-type: none"> Full-thickness anterior supraspinatus tear 	Surgical	<ul style="list-style-type: none"> Arthroscopic rotator cuff repair
7	<ul style="list-style-type: none"> Severe glenohumeral arthritis Severe long head biceps tendinosis Rotator cuff tendinopathy: a low-grade tear of the subscapularis 	Conservative	<ul style="list-style-type: none"> Physical therapy Pain medication Consideration of a corticosteroid injection 	<ul style="list-style-type: none"> Intermediate grade glenohumeral arthritis No significant tendon tears 	Conservative	<ul style="list-style-type: none"> Corticosteroid injection once a year.
8	<ul style="list-style-type: none"> Complete intracapsular rupture of the long head biceps tendon with distal retraction Full-thickness near complete tear of the subscapularis tendon Full-thickness tear of the anterior insertional fibers of the supraspinatus tendon Circumferential superior labrum anteroposterior tear of the glenoid labrum with a paralabral cyst Os acromiale Moderate subcoracoid bursitis. 	Surgical	<ul style="list-style-type: none"> Arthroscopic repair vs débridement Biceps tenodesis Subacromial decompression Labral repair 	<ul style="list-style-type: none"> Full-thickness supraspinatus tear with minimal retraction Upper full-thickness subscapularis tear Effusion at the acromioclavicular joint Meso os acromiale. Some cartilage thinning 	Conservative	<ul style="list-style-type: none"> Physical therapy (since the pain is tolerable)
9	<ul style="list-style-type: none"> Rotator cuff tendinopathy: undersurface tearing of the supraspinatus and infraspinatus Near circumferential tearing of the labrum 	Conservative	<ul style="list-style-type: none"> Physical therapy Pain medication Rest 	<ul style="list-style-type: none"> Frozen shoulder/adhesive capsulitis 	Either	<ul style="list-style-type: none"> Conservative vs surgical Conservative: wait until symptoms resolve

10	<ul style="list-style-type: none"> • Full-thickness tear of the supraspinatus • High-grade interstitial tearing of the subscapularis • Torn and distally retracted long head biceps tendon • Mild glenohumeral arthritis. 	Both	<ul style="list-style-type: none"> • Consideration of a corticosteroid injection • Pain management: NSAIDs/nerve blocks/non-opioid analgesics • Arthroscopic rotator cuff repair with biceps tenodesis • Possible arthroplasty 	<ul style="list-style-type: none"> • Full-thickness supraspinatus tear retracted to the medial aspect of the humeral head • Full-thickness subscapularis tear possibly limited to the upper border • Chronic proximal biceps rupture present • Grade 1 fatty infiltration. • Comminuted proximal humerus fracture that extends to the shaft with some humeral head involvement. 	<ul style="list-style-type: none"> • Surgical: manipulation under anesthesia and capsular release • Arthroscopic rotator cuff repair
11	<ul style="list-style-type: none"> • Proximal humerus fracture 	Surgical	<ul style="list-style-type: none"> • Open reduction internal fixation 	<ul style="list-style-type: none"> • Full-thickness supraspinatus tear retracted to the medial aspect of the humeral head • Full-thickness subscapularis tear possibly limited to the upper border • Chronic proximal biceps rupture present • Grade 1 fatty infiltration. • Comminuted proximal humerus fracture that extends to the shaft with some humeral head involvement. 	<ul style="list-style-type: none"> • Open reduction internal fixation
12	<ul style="list-style-type: none"> • Full-thickness tear of the supraspinatus tendon • Moderate supraspinatus atrophy • Low-grade tear of the cranial fibers of the subscapularis with subluxation of the biceps tendon into the torn subscapularis fibers • Possible anterior-inferior labral tear. 	Surgical	<ul style="list-style-type: none"> • Arthroscopic rotator cuff repair • Possible labral repair 	<ul style="list-style-type: none"> • Full-thickness supraspinatus tear retracted to the medial humeral head with associated grade 2/3 atrophy • Upper border subscapularis tear • Minimal arthritic changes. 	<ul style="list-style-type: none"> • Conservative vs surgical • Conservative: physical therapy and activity modification • Surgical: rotator cuff repair with a subacromial balloon augment • Given her age and the chronicity of her tear, there is a 25% chance that the rotator cuff would fail to heal after the repair
13	<ul style="list-style-type: none"> • Massive rotator cuff tear with full-thickness tears of the distal supraspinatus, infraspinatus, and subscapularis tendons, accompanied by tendon retraction and muscle atrophy indicating chronic tears • Glenohumeral arthritis • Tendinosis, tenosynovitis, and partial tearing of the long head biceps tendon • Advanced acromioclavicular joint osteoarthritis • Rotator cuff tendinosis with partial thickness tear • Subscapularis tendinosis with a tear • Mild acromioclavicular osteoarthritis • Mild glenohumeral arthritis 	Surgical	<ul style="list-style-type: none"> • No specific treatment in the best management plan • Possible options: rotator cuff repair, débridement, subacromial decompression, biceps tenodesis, or arthroplasty 	<ul style="list-style-type: none"> • Massive rotator cuff tear involving the entire supraspinatus, infraspinatus, and subscapularis with stage 4 muscle atrophy • Some early arthritic changes 	<ul style="list-style-type: none"> • Avoid abusive activity • Occasional pain medication • Remain active and exercise
14	<ul style="list-style-type: none"> • Rotator cuff tendinosis with partial thickness tear • Subscapularis tendinosis with a tear • Mild acromioclavicular osteoarthritis • Mild glenohumeral arthritis 	Conservative	<ul style="list-style-type: none"> • Physical therapy • Pain medication • Activity modification 	<ul style="list-style-type: none"> • Full-thickness tear of the supraspinatus with retraction to the medial aspect of the humeral head • Tissue atrophy of the supraspinatus muscle • Modest arthritic changes in cartilage • Mild to moderate acromioclavicular joint separation involving the acromioclavicular and coracoclavicular ligaments 	<ul style="list-style-type: none"> • Conservative vs surgical • Surgical: arthroscopic rotator cuff repair with a subacromial balloon augment
15	<ul style="list-style-type: none"> • Acute to subacute high-grade acromioclavicular joint separation injury (Rockwood V) with complete tears of the acromioclavicular and coracoclavicular ligaments • Elevation of the distal clavicle • High-grade grade 2 strain of the trapezius muscle with an associated intramuscular hematoma 	Conservative	<ul style="list-style-type: none"> • Rest • Activity modification • Pain medication • Physical therapy 	<ul style="list-style-type: none"> • Mild to moderate acromioclavicular joint separation involving the acromioclavicular and coracoclavicular ligaments 	<ul style="list-style-type: none"> • 6 weeks of physical therapy and then a follow-up to determine if an acromioclavicular joint reconstruction is needed
16	<ul style="list-style-type: none"> • Glenohumeral arthritis with inferior osteophyte and glenoid deformity. 	Surgical	<ul style="list-style-type: none"> • Anatomic total shoulder arthroplasty vs reverse shoulder arthroplasty 	<ul style="list-style-type: none"> • Significant glenohumeral arthritis with inferior osteophyte off the humeral head and deformity of the glenoid. • Moderate acromioclavicular joint osteoarthritis • Active distal clavicle osteolysis and bony edema 	<ul style="list-style-type: none"> • Reverse shoulder arthroplasty
17	<ul style="list-style-type: none"> • Active distal clavicular osteolysis with moderate supraspinatus tendinosis and an interstitial tear at the footprint 	Conservative	<ul style="list-style-type: none"> • Physical therapy • Activity modification • Pain medication • Consideration of a corticosteroid injection 	<ul style="list-style-type: none"> • Moderate acromioclavicular joint osteoarthritis • Active distal clavicle osteolysis and bony edema 	<ul style="list-style-type: none"> • Conservative vs surgical • Surgical: distal clavicle resection
18	<ul style="list-style-type: none"> • Near full-thickness tear of the supraspinatus tendon • Partial interstitial tearing of the infraspinatus tendon 	Conservative	<ul style="list-style-type: none"> • Physical therapy • Activity modification 	<ul style="list-style-type: none"> • Focal high-grade partial leading edge supraspinatus tear 	<ul style="list-style-type: none"> • Physical therapy • Use arm as tolerated

(continued on next page)

Table 1 (continued)

Patient #	ChatGPT diagnosis	ChatGPT plan (conservative, surgical, or both)	ChatGPT detailed plan	Specialist diagnosis	Specialist plan (conservative surgical, or either)	Specialist detailed plan
	<ul style="list-style-type: none"> Trace subacromial/subdeltoid bursitis 		<ul style="list-style-type: none"> Pain medication Consideration of a corticosteroid injection 			
19	<ul style="list-style-type: none"> Rotator cuff tendinosis Tenosynovitis of the long head biceps tendon. 	Conservative	<ul style="list-style-type: none"> Rest NSAIDs Physical therapy 	<ul style="list-style-type: none"> Intact rotator cuff Healthy looking shoulder 	Conservative	<ul style="list-style-type: none"> Physical therapy Use arm as tolerated
20	<ul style="list-style-type: none"> Severe glenohumeral arthritis 	Conservative	<ul style="list-style-type: none"> Physical therapy Pain medication Activity modification 	<ul style="list-style-type: none"> Severe advanced glenohumeral arthritis with essentially joint space aberration and some posterior subluxation Glenoid shows some moderate posterior wear Glenohumeral internal rotation deficit Some mild rotator interval inflammation No labral tear Rotator cuff is intact Large superior cuff tear No significant glenohumeral arthritis Severe atrophy of the muscle bellies of the supraspinatus and infraspinatus Subscapularis is intact Full-thickness supraspinatus and infraspinatus tear Upper border subscapularis tear Grade 3 fatty atrophy of the supraspinatus and infraspinatus Chondral thinning of the glenohumeral joint 	Surgical	<ul style="list-style-type: none"> Reverse shoulder arthroplasty
21	<ul style="list-style-type: none"> Possible tear of the posterosuperior labrum 	Conservative	<ul style="list-style-type: none"> Physical therapy Rest Activity modification 		Conservative	<ul style="list-style-type: none"> Physical therapy
22	<ul style="list-style-type: none"> Massive rotator cuff tear Glenohumeral arthritis Acromioclavicular joint arthritis 	Conservative	<ul style="list-style-type: none"> Monitor symptoms 		Conservative	<ul style="list-style-type: none"> Wait to see how long the cortisone injection works for and if the pain worsens.
23	<ul style="list-style-type: none"> Full-thickness tear of the supraspinatus Moderate fatty atrophy of the supraspinatus Mild fatty atrophy of the infraspinatus Complete tear of the long head of the biceps tendon Moderate joint effusion with decompression into the subacromial/subdeltoid bursa Os acromiale with degenerative changes across the synchondrosis 	Conservative	<ul style="list-style-type: none"> Physical therapy NSAIDs Consideration of a corticosteroid injection 		Surgical	<ul style="list-style-type: none"> Reverse shoulder arthroplasty
24	<ul style="list-style-type: none"> High-grade partial-thickness tearing involving the posterior supraspinatus and infraspinatus tendons Posterior labral tear with accompanying paralabral cyst 	Conservative	<ul style="list-style-type: none"> Physical therapy NSAIDs Activity modification Consideration of a corticosteroid injection 	<ul style="list-style-type: none"> Focal tearing in the area of the supraspinatus, which appears to be essentially full-thickness in nature with minimal retraction The muscle bellies of the supraspinatus are good A posterior inferior paralabral cyst Cartilage surfaces appeared to be generally well preserved No advanced glenohumeral arthritis. Mild case of adhesive capsulitis (frozen shoulder) Evidence of some inflammation No rotator cuff tearing No arthritis 	Conservative	<ul style="list-style-type: none"> Physical therapy
25	<ul style="list-style-type: none"> Adhesive capsulitis Mild acromioclavicular osteoarthritis Mild infraspinatus tendinosis 	Conservative	<ul style="list-style-type: none"> Physical therapy Pain medication Consideration of a corticosteroid injection 		Conservative	<ul style="list-style-type: none"> Possible options: benign neglect, corticosteroid injection, physical therapy

26	<ul style="list-style-type: none">• Nondisplaced fracture of the greater tuberosity of the humerus	Conservative	<ul style="list-style-type: none">• Physical therapy• Pain medication• Activity modification	<ul style="list-style-type: none">• A relatively nondisplaced fracture of the greater tuberosity• A large amount of stress shielding distally	Conservative	<ul style="list-style-type: none">• Remain in sling for the next 2 weeks. Come out of the sling to range her elbow, wrist, and hand• Start physical therapy after the 2 weeks.
27	<ul style="list-style-type: none">• Glenohumeral arthritis	Both	<ul style="list-style-type: none">• Consideration of a corticosteroid injection• Referral to an orthopedic specialist	<ul style="list-style-type: none">• Components are well reduced.• Severe glenohumeral arthritis with joint space narrowing, some proximal humeral migration, and osteophytes of the acromion and humeral head.	Surgical	<ul style="list-style-type: none">• Reverse shoulder arthroplasty
28	<ul style="list-style-type: none">• Glenohumeral subluxation	humeral posterior	<ul style="list-style-type: none">• Arthroscopic débridement anatomic total shoulder arthroplasty	<ul style="list-style-type: none">• Glenohumeral arthritis vs posterior humeral subluxation• Erosive changes with loss of joint space and eccentric wear of the glenoid posteriorly	with Surgical	<ul style="list-style-type: none">• Reverse shoulder arthroplasty
29	<ul style="list-style-type: none">• Glenohumeral arthritis	Surgical	<ul style="list-style-type: none">• Consultation with an orthopedic surgeon• Possible options: arthroscopic débridement, joint fusion, anatomic total shoulder arthroplasty, reverse shoulder arthroplasty	<ul style="list-style-type: none">• Glenohumeral arthritis• The humeral head is centered on the glenoid fossa• Mild medialization of the joint line	Surgical	<ul style="list-style-type: none">• Reverse shoulder arthroplasty

NSAIDs, non-steroidal anti-inflammatory drugs.

Table II
Characteristics of involved patients and performance of ChatGPT in our study.

	N	%
Gender		
Males	15	52
Females	14	48
Patient diagnosis		
Rotator cuff repair	12	41.0
Mild inflammation	2	6.9
Acromio-clavicular joint trauma	2	6.9
Adhesive capsulitis	2	6.9
Osteoarthritis	8	28
Proximal humerus fracture	3	10.3
Performance		
Missed diagnosis	2	6.9
Missed management	5	17.2
Correct but insufficient management plan	6	20.7
Provided one but not all valid management plans	5	17.2

osteoarthritis, and 33% of proximal humeral fractures. Moreover, of the remaining 24 patients that were managed correctly, ChatGPT did not specify the appropriate management in 6 patients (4 patients that should be managed surgically, and where ChatGPT picked surgical management without specifying the surgical management, and 2 patients where surgical management was most appropriate, both options of surgical and conservative management were given by the AI model) (Table II). In addition, ChatGPT chose only one management (conservative or surgical) in 5 patients, where both were applicable and dependent on the patient's choice. Therefore, it can be said that ChatGPT showed poor management in 55% of the cases.

Discussion

To the best of our knowledge, this is the first study to evaluate the diagnostic accuracy of ChatGPT in the clinical setting for common complaints of the shoulder and the elbow. ChatGPT has been demonstrated to be extremely useful in a clinical context, creating effective discharge summaries, predicting disease risk and outcome, and identifying the imaging steps required for breast cancer screening and the assessment of breast pain.^{13,17,20,21} By making accessible and understandable health information available to the general population, ChatGPT in healthcare settings also has the potential to increase health literacy and advance customized medicine.^{1,4,6,15,27} However, this is not without limitations, such as the possibility of bias and transparency, healthcare cybersecurity, and data privacy, which would raise ethical questions.^{4,13,18,20,21} Furthermore, the creation of erroneous content can have serious detrimental effects on health care; as a result, this legitimate issue should be carefully taken into account in clinical practice.^{1,2,4,20} This worry also extends to ChatGPT's capacity to offer explanations for erroneous judgments.²¹ A careful consideration should also be given to medicolegal and accountability problems in the event that the ChatGPT application causes medical errors.³ Importantly, ChatGPT is unable to understand biologic system complexity, a crucial idea required for research and decision-making in the field of health care.¹⁷

The AI model was able to correctly choose the diagnosis in 93% (27/29) of patients. This rate is similar to another study that examined the accuracy of ChatGPT in giving a list of 10 differential diagnoses for common chief complaints.¹² In fact, the patients misdiagnosed by ChatGPT had mild inflammation in an intact shoulder. This can be explained by many factors. First, ChatGPT is very dependent on the magnetic resonance imaging report, therefore exaggerating some of the diagnoses seen in many of the patients from this cohort and thus leading to a misdiagnosis and a potential mismanagement. Second,

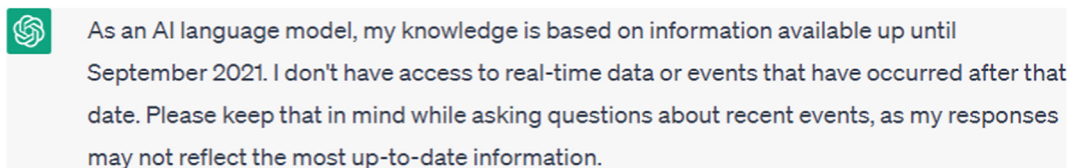


Figure 2 A statement of ChatGPT showing its lack of knowledge after September 2021.

the patient might leave something out (such as psychosocial or other issues) that a specialist might remember to ask and this information may change the whole diagnosis/management. This is not possible with ChatGPT because there is no interaction between the patient and the AI. Third, a risk that was not seen in the study due to the involvement of medical professionals creates a difficulty for the general user to input an optimal medical term that may eventually lead to misinformation and an erroneous diagnosis.

The rate of mismanagement was higher reaching 17.2% (5/29). In fact, ChatGPT mismanaged patients presenting with rotator cuff injuries, osteoarthritis, proximal humeral fractures which constitute some of the most common presenting symptoms a specialist may confront. Furthermore, even when managed correctly, the AI tool was not able to specify whether to go for conservative or surgical treatment in some patients. This can be explained by the tool's lack of interaction with the patient as well as the lack of empathy which would underestimate the patient's pain and loss of function. In fact, this lack of interaction with the patient and history taking was shown to be one of the most important factors in deciding on surgery.²⁵ Furthermore, Gupta et al showed in a recent study that surgeon-specific factors which are not taken into account by the AI were more important in deciding for surgery than patient-specific factors.¹⁰ Another drawback is that when a surgical or conservative treatment plan was chosen, the detailed plan was lacking in some cases (such as proposing both anatomic and reverse total shoulder arthroplasty, or open reduction internal fixation versus shoulder arthroplasty for fracture, etc.). In addition, in some patients where both conservative and surgical management were applicable, ChatGPT chose only one. This can as well be explained by the lack of interaction and empathy of the AI which is well-documented in the literature.^{1,11}

The issue of undervaluing the role of the human brain in health care practice and research should not be disregarded, so it is crucial to address any psychologic, financial, and social ramifications that might follow the use of ChatGPT tools in healthcare settings. Other limitations of ChatGPT include the fact that it can be biased and lacking recent knowledge.²² In fact, racial, religious, and gender biases have been reported.¹² Furthermore, since ChatGPT is not connected to the internet, it lacks any knowledge beyond September 2021 (Fig. 2) which renders this tool inapplicable in cases where new surgical techniques and controversies are thriving such as in the management of massive irreparable rotator cuff tears where multiple treatment options continue to evolve.

Strengths and limitations

The use of ChatGPT-3, whose user interface is comfortable for ordinary users due to its free text writing style and text style output, is one of this study's strengths. Additionally, rather than using vignettes, this study was based on specific cases of actual patients. However, some of the limitations included the limited number of included patients which makes this a pilot study and the fact that there was no third-party observer to see whether the specialist had the right diagnosis and management. Furthermore,

diagnosis and management of ChatGPT was compared to that of an experienced fellowship-trained shoulder surgeon, and perhaps using more than one surgeon may have provided more reliable results. However, considering the cases were not controversial, we believe our results would not have differed had more than one surgeon been involved in diagnosis and management. Considering that this study is the first of its kind, we believe it holds valuable merit and would serve as a foundation for future studies that can involve a higher number of surgeons and more complex shoulder cases.

Conclusion

Although ChatGPT was a worthy opponent, in its current form it will not be able to fully replace a shoulder and elbow specialist in diagnosing and treating patients for many reasons, such as misdiagnosis in 7% of cases, poor management in 55% of cases, lack of a clear management strategy, and failure to consider the patient's perspective or the surgeon's expertise. Furthermore, the magnetic resonance imaging report, which has a tendency to exaggerate findings, had a significant impact on ChatGPT's choice. Moreover, the lack of the AI's knowledge after September 2021 makes its usage inapplicable in cases of recent surgical techniques where controversies are present.

Disclaimers:

Funding: No funding was disclosed by the authors.

Conflicts of interest: The authors, their immediate families, and any research foundation with which they are affiliated have not received any financial payments or other benefits from any commercial entity related to the subject of this article.

References

1. Ahn C. Exploring ChatGPT for information of cardiopulmonary resuscitation. *Resuscitation* 2023;185:109729. <https://doi.org/10.1016/j.resuscitation.2023.109729>.
2. Ali SR, Dobbs TD, Hutchings HA, Whitaker IS. Using ChatGPT to write patient clinic letters. *Lancet Digit Health* 2023;5:e179–81. [https://doi.org/10.1016/S2589-7500\(23\)00048-1](https://doi.org/10.1016/S2589-7500(23)00048-1).
3. Biswas S. ChatGPT and the future of medical writing. *Radiology* 2023;307, e223312. <https://doi.org/10.1148/radiol.223312>.
4. D'Amico RS, White TG, Shah HA, Langer DJ. I asked a ChatGPT to write an editorial about how we can incorporate chatbots into neurosurgical research and patient care.... *Neurosurgery* 2023;92:663–4. <https://doi.org/10.1227/neu.0000000000002414>.
5. Daher M, Ghanimeh J, Otayek J, Ghoul A, Bizdikian A-J, EL Abiad R. Augmented reality and shoulder replacement: a state of the art review article. *JSSES Rev Rep Tech* 2023;3:274–8. <https://doi.org/10.1016/j.xrrt.2023.01.008>.
6. Duong D, Solomon BD. Analysis of large-language model versus human performance for genetics questions. *medRxiv* 2023. <https://doi.org/10.1101/2023.01.27.23285115>.
7. Else H. Abstracts written by ChatGPT fool scientists. *Nature* 2023;613:423. <https://doi.org/10.1038/d41586-023-00056-7>.
8. 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 Med Educ* 2023;9:e45312. <https://doi.org/10.2196/45312>.
9. Goodman RS, Patrinely JR, Osterman T, Wheless L, Johnson DB. On the cusp: considering the impact of artificial intelligence language models in healthcare. *Med* 2023;4:139–40. <https://doi.org/10.1016/j.medj.2023.02.008>.

10. Gupta R, Jung J, Johnston TR, Wright DJ, Uong J, Lim PK, et al. Surgeon-specific factors have a larger impact on decision-making for the management of proximal humerus fractures than patient-specific factors: a prospective cohort study. *J Shoulder Elbow Surg* 2023;32:1701–9. <https://doi.org/10.1016/j.jse.2022.12.016>.
11. Hallsworth JE, Udaondo Z, Pedrós-Alió C, Höfer J, Benison KC, Lloyd KG, et al. Scientific novelty beyond the experiment. *Microb Biotechnol* 2023;16:1131–73. <https://doi.org/10.1111/1751-7915.14222>.
12. Hirosawa T, Harada Y, Yokose M, Sakamoto T, Kawamura R, Shimizu T. Diagnostic accuracy of differential-diagnosis lists generated by generative pre-trained transformer 3 chatbot for clinical vignettes with common chief complaints: a pilot study. *Int J Environ Res Public Health* 2023;20:3378. <https://doi.org/10.3390/ijerph20043378>.
13. Holzinger A, Keiblinger K, Holub P, Zatloukal K, Müller H. AI for life: trends in artificial intelligence for biotechnology. *N Biotechnol* 2023;74:16–24. <https://doi.org/10.1016/j.nbt.2023.02.001>.
14. Huh S. Are ChatGPT's knowledge and interpretation ability comparable to those of medical students in Korea for taking a parasitology examination?: a descriptive study. *J Educ Eval Health Prof* 2023;20:1. <https://doi.org/10.3352/jeehp.2023.20.1>.
15. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLOS Digit Health* 2023;2:e0000198. <https://doi.org/10.1371/journal.pdig.0000198>.
16. Liebreiz M, Schleifer R, Buadze A, Bhugra D, Smith A. Generating scholarly content with ChatGPT: ethical challenges for medical publishing. *Lancet Digit Health* 2023;5:e105–6. [https://doi.org/10.1016/S2589-7500\(23\)00019-5](https://doi.org/10.1016/S2589-7500(23)00019-5).
17. Mann DL. Artificial intelligence discusses the role of artificial intelligence in translational medicine: a JACC: basic to translational science interview with ChatGPT. *JACC Basic Transl Sci* 2023;8:221–3. <https://doi.org/10.1016/j.jacbs.2023.01.001>.
18. Mijwil M, Aljanabi M, Ali AH. ChatGPT: exploring the role of cybersecurity in the protection of medical information. *Mesopotamian J Cyber Secur* 2023;18–21. <https://doi.org/10.58496/MJCS/2023/004>.
19. Open AI. Introducing ChatGPT. Available at: <https://openai.com/blog/chatgpt/>. Accessed April 1, 2023.
20. Patel SB, Lam K. ChatGPT: the future of discharge summaries? *Lancet Digit Health* 2023;5:e107–8. [https://doi.org/10.1016/S2589-7500\(23\)00021-3](https://doi.org/10.1016/S2589-7500(23)00021-3).
21. Rao A, Kim J, Kamineni M, Pang M, Lie W, Succi MD. Evaluating ChatGPT as an adjunct for radiologic decision-making. *medRxiv* 2023. <https://doi.org/10.1101/2023.02.02.23285399>.
22. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare (Basel)* 2023;11:887. <https://doi.org/10.3390/healthcare11060887>.
23. Sezgin E, Sirrianni J, Linwood SL. Operationalizing and implementing pre-trained, large artificial intelligence linguistic models in the US health care system: outlook of generative pretrained transformer 3 (GPT-3) as a service model. *JMIR Med Inform* 2022;10:e32875. <https://doi.org/10.2196/32875>.
24. SimilarWeb. ChatGPT tops 25 million daily visits. Available at: <https://www.similarweb.com/blog/insights/ai-news/chatgpt-25-million/>. Accessed April 1, 2023.
25. Simon MJ, Regan WD. Utilization of MRI in surgical decision making in the shoulder. *BMC Musculoskelet Disord* 2022;23:588. <https://doi.org/10.1186/s12891-022-05541-0>.
26. Sinha RK, Deb Roy A, Kumar N, Mondal H. Applicability of ChatGPT in assisting to solve higher order problems in pathology. *Cureus* 2023;15:e35237. <https://doi.org/10.7759/cureus.35237>.
27. Yeo YH, Samaan JS, Ng WH, Ting P-S, Trivedi H, Vipani A, et al. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. *Clin Mol Hepatol* 2023;29:721–32. <https://doi.org/10.3350/cmh.2023.0089>.