information

MDPI

Article

Descriptive Heterogeneity of the Hornblower Sign Across Scientific Literature, Search Engines, and Large Language Models

Peter Melcher [1], Ralf Henkelmann [2], Susanne Schleif [3] and Salim Youssef [2,*]

[1] Department of Orthopedics and Trauma Surgery, Helios Klinik Leisnig, 04703 Leisnig, Germany; peter.melcher@helios-gesundheit.de
[2] Department of Orthopedics, Trauma and Plastic Surgery, University of Leipzig, 04103 Leipzig, Germany; ralf.henkelmann@medizin.uni-leipzig.de
[3] Independent Researcher, 04109 Leipzig, Germany
* Correspondence: salim.youssef@medizin.uni-leipzig.de

Abstract

Background/Objectives: Digitalization of medical knowledge has improved access to information but also increased the spread of imprecise content. Repeated exposure to incorrect descriptions may lead to their normalization over time. This is particularly evident for the Hornblower sign, which is frequently conflated with the Patte test in the literature, online sources, and large language model outputs. This study systematically evaluates these descriptions and quantifies related inaccuracies. **Methods**: A three-step approach was applied to answer the question "What is the Hornblower sign?". First, primary publications referring to the original description by Arthuis were analyzed. Second, the first 35 Google search results were systematically reviewed. Third, responses from five widely used LLMs (ChatGPT 5.1, Grok 4.1, Gemini 3 Pro, Perplexity, and DeepSeek-R1) were evaluated. All descriptions were assessed using a standardized 4-point scoring system (0–3 points) capturing content accuracy and correct differentiation between the Hornblower sign and the Patte test. **Results**: Fourteen original publications were included, yielding a mean score of 2.07. Correct descriptions were found in 50%, while 43% described only the Patte test. Among 34 evaluable Google search results, the mean score was 1.17, with 77% scoring ≤ 1 point. The five LLMs achieved a mean score of 1.8, demonstrating substantial variability and overall incomplete conceptual accuracy. **Conclusions**: Descriptions of the Hornblower sign show substantial heterogeneity and frequent inaccuracies across the scientific literature, online sources, and LLM outputs. Conflation with the Patte test undermines diagnostic reliability and limits study comparability. Critical source appraisal and adherence to original test descriptions are essential to maintain clinical and scientific rigor.

Keywords: Hornblower sign; Patte test; diagnostic tests; medical misinformation; large language models; bias

Academic Editor: Giorgio Maria Di Nunzio

Received: 8 January 2026
Revised: 5 February 2026
Accepted: 3 March 2026
Published: 5 March 2026

1. Introduction

The increasing digitalization of medical knowledge resources has substantially improved access to subject-specific information [1,2]. Search engines, social media platforms, and, more recently, AI-based text generation systems are now widely used for the retrieval, preparation, and dissemination of medical information [1–3]. Despite their efficiency, however, these technologies have a limited capacity to assess the reliability and epistemic status of the information they reproduce [4,5].

In the context of physical examination, accurate and standardized descriptions of diagnostic tests are essential prerequisites for reliable clinical decision-making and for the interpretation of research findings [6,7]. Deviations from original test definitions may lead to heterogeneous execution and inconsistent diagnostic conclusions. Such distortions are particularly problematic when they concern established clinical signs that are widely cited and taught, as repeated reproduction of imprecise descriptions may normalize inaccuracies over time [8–10].

The Hornblower sign represents a paradigmatic example of this phenomenon. Originally described by Arthuis in 1972, the sign refers to a compensatory abduction maneuver required to bring the hand to the mouth in the presence of external rotator paralysis [11]. In contemporary orthopedic literature, however, the Hornblower sign is frequently conflated with the Patte test [12–14]. Despite superficial similarities, the two maneuvers differ biomechanically and are associated with distinct patterns of muscular activation [15,16]. Failure to distinguish between them may therefore affect diagnostic interpretation and limit the comparability of clinical studies.

Although inaccuracies in the description of clinical examination techniques have been reported in orthopedic literature [17,18], the extent to which a specific examination maneuver is inconsistently defined across different information ecosystems remains unclear. In particular, no systematic analysis has quantified how descriptions of the Hornblower sign deviate from its original definition across primary scientific literature, publicly accessible online sources retrieved via search engines, and contemporary LLM outputs.

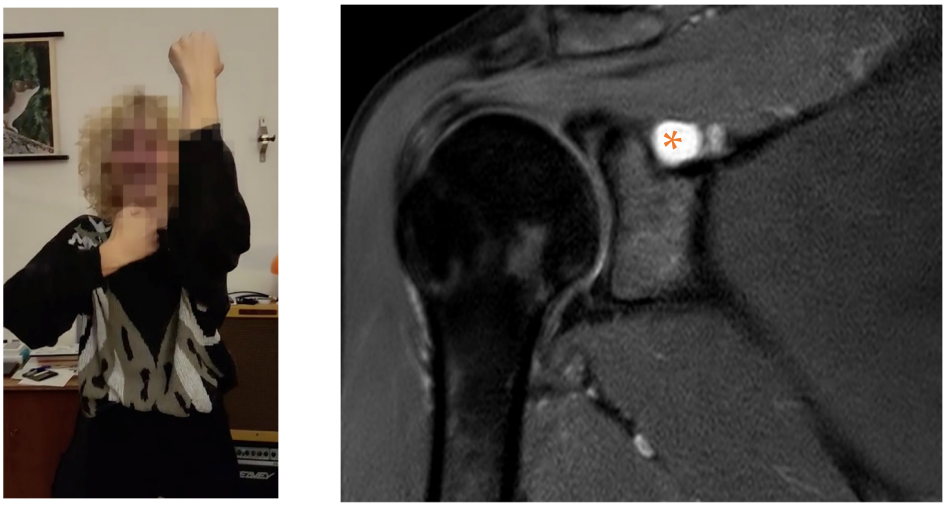
Therefore, the objective of this study was to systematically evaluate and compare descriptions of the Hornblower sign across these three source types. By quantifying content accuracy and the degree of conflation with the Patte test, this study aims to highlight potential sources of diagnostic ambiguity and to illustrate broader challenges in the digital dissemination of medical knowledge.

2. Materials and Methods

2.1. Defining the Hornblower Sign

The Hornblower sign was originally described by Arthuis in 1972 in patients with obstetric brachial plexus palsy. In this condition, paralysis predominantly affects the shoulder's external rotators, resulting in a functional predominance of the internal rotators [11]. As a late consequence, patients may develop the so-called "clairon" or Hornblower sign, in which compensatory abduction of the arm is required to bring the hand to the mouth. A representative example of a positive Hornblower sign is shown in Figure 1a. Notably, Arthuis exclusively referred to obstetric paralysis and did not describe a specific test performed in high external rotation. On the contrary, he explicitly warned against this arm position, as it may facilitate posterior shoulder dislocation in the presence of such paralysis.

In an orthopedic context, the Hornblower sign provides clinical evidence of dysfunction of the teres minor or infraspinatus muscles, which are the primary external rotators of the shoulder [19]. In several publications, a Hornblower sign has been described as a functional assessment performed in high external rotation (90° abduction, 90° elbow flexion, external rotation against resistance) [12,13]. However, this maneuver was originally described by Patte and should not be equated with the Hornblower sign [14]. Despite superficial similarities, the two tests differ biomechanically and may indicate distinct underlying pathologies.



(**a**) (**b**)

Figure 1. Patient presenting with a positive Hornblower sign but a negative Patte test: (**a**) Positive Hornblower sign; (**b**) magnetic resonance imaging (MRI) depicting spinoglenoidal cyst (*), responsible for isolated infraspinatus muscle insufficiency.

The Hornblower sign is more sensitive and specific for infraspinatus dysfunction or global external rotator cuff insufficiency, whereas the Patte test is more sensitive and specific for isolated teres minor dysfunction [20]. The clinical relevance of this distinction is illustrated by a patient presenting with a positive Hornblower sign (Figure 1a) but a negative Patte test. In this case, a traumatic injury led to the formation of a spinoglenoidal cyst (Figure 1b), resulting in compression of the suprascapular nerve and subsequent isolated insufficiency of the infraspinatus muscle.

2.2. Systematic Analysis

For the present study, a systematic analysis of various sources describing the Hornblower sign was conducted. The aim was to comprehensively characterize the range and nature of potential misdescriptions of this clinical test across scientific literature, publicly available sources, and AI-based platforms.

The analysis was based on a three-tiered source strategy to address the research question, "What is the Hornblower sign?":

(1) Primary studies citing the original publication by Arthuis.
(2) Results obtained from a Google search using the identical query.
(3) Responses generated by different LLMs to the same question.

2.2.1. Primary Studies

Within the group of original publications, all 24 citations of the original article by Arthuis listed in Google Scholar were evaluated. Sources that merely cited the original work without providing independent substantive content were excluded (n = 5), as were duplicate entries (n = 3) and publications lacking a concrete description of the Hornblower sign (n = 2). In total, 14 original publications were included in the final analysis.

2.2.2. Google Search

As part of the Google search, the first 35 search results were evaluated. To minimize personalization effects, the search was conducted in the browser's incognito mode, without an active user login and with no access to prior search history or stored cookies. Sources without thematic relevance to the Hornblower sign were excluded (n = 1). When con-

tent was available from the same publisher across multiple formats, the publisher was considered only once (n = 2); in cases of differing scores for the same publisher, the most frequently assigned score was used. The analysis ultimately included textual, image-based, and video descriptions from a total of 30 distinct publishers.

2.2.3. Large Language Models

Previous studies have demonstrated that LLMs can provide generally accurate responses to medical questions. At the same time, relevant differences in content accuracy between individual models have been reported [21]; therefore, the analysis deliberately did not focus on a single model. Hence the above research question was posed in parallel to five different LLMs on 24 November 2025: ChatGPT 5.1 (OpenAI, San Francisco, CA, USA); Grok 4.1 (xAI, San Francisco, CA, USA), Gemini 3 Pro (Google DeepMind, London, UK), Comet Assistant (Perplexity, San Francisco, CA, USA, https://www.perplexity.ai/), and DeepSeek-R1 (DeepSeek, Hangzhou, China). The models were selected based on their widespread use, accessibility, and relevance in both academic and public contexts [22]. All LLM queries were conducted as single-prompt requests, with each query performed in a separate conversation thread to minimize potential bias from prior interactions and to reflect a typical single-interaction user experience rather than intra-model response variability [3,23].

2.2.4. Scoring System

The evaluation of test descriptions was performed using a predefined scoring system specifically developed to assess the accuracy and clinical reliability of descriptions of the Hornblower sign. A four-point ordinal scale (0–3) was deliberately chosen to reflect clinically meaningful differences in test interpretation:

3 points: Correct description or depiction of the Hornblower sign.
2 points: Description of the Patte test (or lag sign in high external rotation) labeled as the Hornblower sign, while simultaneously presenting the correct Hornblower sign as an alternative test procedure.
1 point: Exclusive description of the Patte test (or lag sign in high external rotation).
0 points: Incorrect description of the Hornblower sign in which a pathological test result could be interpreted as negative, and thus as normal, even if the Patte test is correctly described.

The primary rationale of this approach was to capture the suitability of a given description for routine clinical practice and its comparability across scientific studies. All sources were independently reviewed and scored by two reviewers. In cases of discrepant ratings or uncertainty, the respective source was re-evaluated jointly, and a consensus score was determined through discussion.

3. Results

3.1. Primary Studies

In the group of studies referencing Arthuis' original study, seven of fourteen (50%) correctly described the Hornblower sign. One study (7%) included an image depicting a positive Hornblower sign in addition to a correct description of the Patte test and therefore received 2 points. Six studies (43%) described only the Patte test. No incorrect test descriptions were identified. The mean score within this group was 2.07 (Table 1).

Table 1. Studies citing Arthuis' original description of the Hornblower sign.

References	Points
Bakhsh and Nicandri [24]	1
Gerber et al. [12]	1
Walch et al. [19]	2
Williams et al. [13]	1
Panayiotou Charalambous [25]	3
Codsi et al. [26]	1
Kim et al. [27]	3
Bi et al. [28]	1
Razmjou [29]	3
Razmjou [30]	3
Kim et al. [31]	3
Pandey [32]	3
Robinson et al. [33]	1
Blønd [34]	3
Mean	2.07

3.2. Google Search

In total, 34 search results from 30 publishers from heterogeneous online sources were included in the analysis. The following score distribution was identified: 3 sources received 3 points (10%), 4 sources received 2 points (13%), 18 sources received 1 point (60%), and 5 sources received 0 points (17%). The mean score was 1.17 (Table 2). Among the Google search results, mean scores were comparable between sources that included video material (n = 21; mean score, 1.26) and those without video content (n = 13; mean score, 1.14) (Table 2).

3.3. Large Language Models

The evaluation of the LLM responses showed a mean score of 1.8 points, with individual scores of 1 point for both ChatGPT 5.1 and Gemini Pro (n = 2, 40%), 2 points for both Perplexity and DeepSeek (n = 2, 40%), and 3 points for Grok 4.1 (n = 1, 20%). No misinterpretations of a positive test as a normal finding were observed (Table 3). Among the evaluated large language models, response length varied substantially, ranging from 152 to 646 words. No apparent correlation between response length and accuracy score was identified. Responses with comparable word counts yielded markedly different scores, and both shorter and longer responses were observed across intermediate score levels (Table 3).

3.4. Comparison of Source Types

When comparing source types, primary scientific literature achieved the highest mean score (2.07; 95% CI, 1.50–2.65; n = 14), followed by large language models (1.8; 95% CI, 0.76–2.84; n = 5), whereas Google search results showed the lowest mean score (1.17; 95% CI, 0.86–1.48; n = 30).

Table 2. Results of the Google search, including scores and media types; X indicates the frequency of occurrence of a given source in the search results (X = once, XX = twice, XXX = three times).

Source	Points	Website	Study	FaceBook	Instagram	YouTube	X.com	PDF	Reddit	Includes Video
Physiopedia	2	X		X		X				XXX
DocCheck	1	X								
Physiotutors	1	XX		X						XXX
Matassessment	1	X								X
Journal of Orthopaedic Science	3		X							
Journal of Bone & Joint Surgery	2		X							
Cirugía de Hombro y Codo	3			X						X
OrthoFixar—Orthopedic Surgery	2			X						
Physical Therapie Haven	1	X								X
dr_priyanka_physical_activity_	0				X					X
wikism	1	X								X
Orthobullets	1	X								X
Cirugía de Hombro y Codo—Guido Fierro MD	3					X				X
Physical Therapy Web	0	X								X
matt_pt_dip_mdt	1				X					X
Ccedseminars	1					X				X
Ortho Eval Pal with Paul Marquis PT	1					X				X
Sports Med Review	0								X	X
Physiotherapy Online	0	X								
Dr. Gerald Diaz	1	X								X
mindluster	1	X								X
Elsevier	1	X								
CRTechnologies	2					X				X
The Clinical Orthopaedic Society	1	X								
Bangalore Shoulder Institute	1	X								
Carepatron	1	X								
physio_clinics	1						X			
shoulderdoc	1	X								
Abteilung Orthopädie und Sporttraumatologie: Klink am Ring	0							X		
Medical Dictionary	1	X								
		18	2	4	2	5	1	1	1	
	Mean 1.17	Total 34								21

Table 3. Results of the LLM responses including scores and word count.

	ChatGPT 5.1	Grok 4.1	Gemini Pro	Perplexity	DeepSeek	Mean
Points	1	3	1	2	2	1.8
Word Count	176	283	283	152	646	308

4. Discussion

In response to the study objective of evaluating how accurately the Hornblower sign is described across scientific literature, online sources, and large language models, the present findings demonstrate that inaccuracies in medical test descriptions are not confined to informal educational material but are also embedded within peer-reviewed orthopedic research. Across source types, Google search results exhibited the lowest overall accuracy, while large language models achieved intermediate scores.

4.1. Mechanisms Within Scientific Literature

The observed inconsistencies cannot be attributed solely to informal or non-academic sources. Notably, even among studies directly referencing Arthuis' original work, only half provided a correct description of the Hornblower sign. This observation aligns with prior analyses demonstrating that citation of a primary source does not necessarily imply conceptual fidelity to its content [35,36]. Once a modified or inaccurate interpretation of a clinical test enters the literature, it may be perpetuated through secondary citation rather than direct engagement with the original publication, thereby contributing to misinterpretation and the continued dissemination of imprecise or incorrect descriptions [35,36]. In the specific case of the Hornblower sign, reliance on secondary sources may have been exacerbated by the limited accessibility of Arthuis' original 1972 publication, which is neither freely available nor published in languages other than French [11].

In addition, information is more readily accepted without questioning when it is presented by recognized authorities, a phenomenon commonly referred to as authority bias [37]. In this context, incorrect or incomplete descriptions of the Hornblower sign presented in high-impact journals or by recognized experts may be taken for granted and subsequently propagated, shifting emphasis from whether a statement is correct to *who* articulated it or *where* it was published. Once such inaccuracies become institutionalized, repeated rewording and reiteration in secondary sources can progressively distort the original information, thereby reinforcing existing interpretations and limiting critical reassessment [35,38].

4.2. Digital Amplification Through Search Engines and Online Media

Digital dissemination mechanisms further amplify these processes. Search engines prioritize visibility based on algorithmic criteria such as keyword density, indexing, and user engagement rather than content accuracy, which may favor the dominance of frequently reproduced but potentially inaccurate content [39]. The low mean score observed among Google search results in this study (mean score = 1.17) reflects this dynamic, suggesting a systematic drift away from the original test definition in publicly accessible online content.

The widespread use of video-based educational material may further contribute to this effect. Video formats often favor simplified, visually intuitive demonstrations that may not adequately convey subtle but clinically relevant distinctions between examination techniques [40]. In the present analysis, sources including video material (mean score = 1.26) did not demonstrate clearly superior descriptive accuracy compared with no-video sources

(mean score = 1.14), indicating that multimedia presentation alone does not ensure higher informational quality.

4.3. Positioning of Large Language Models

Large language models appear to integrate seamlessly into this existing information ecosystem. The intermediate accuracy observed across all evaluated LLMs (mean score = 1.8) is consistent with prior literature indicating that, rather than correcting long-standing inconsistencies, these models predominantly reflect the dominant descriptions present in their training data and readily accessible sources [41]. This tendency is compatible with the probabilistic nature of LLMs, which are optimized to generate plausible and contextually coherent responses rather than to verify epistemic correctness against primary sources [4,5]. As a result, misconceptions that are already well established within scientific and digital sources may persist in model-generated outputs and be further amplified through repeated use.

4.4. Clinical and Methodological Implications

From a clinical perspective, inaccurate or inconsistent descriptions of the Hornblower sign may lead to erroneous test performance and misinterpretation of findings, potentially affecting diagnostic reasoning in the assessment of posterosuperior rotator cuff pathology. The frequent conflation of the Hornblower sign with the Patte test observed in this study (Primary Studies = 50%, Google-Search = 73%; LLM = 80%) is particularly problematic, as the two maneuvers differ in execution and underlying muscular demands despite both being described as tests of external rotation. Experimental work has demonstrated that shoulder external rotation is highly position-dependent, with both inter- and intramuscular differences in recruitment patterns [15,16]. In particular, external rotation at 90° of abduction preferentially engages the teres minor, whereas lower degrees of abduction rely more strongly on the infraspinatus [20,42]. Clinical cases such as the one presented in this study, in which the Hornblower sign was positive while the Patte test was negative, highlight the importance of differentiating between the two tests. Given the low prevalence of isolated infraspinatus rupture, reported at approximately 2% in surgically treated cohorts [43], clinicians are rarely exposed to truly positive Hornblower signs, which may contribute to the persistence of inaccurate test descriptions due to limited opportunities for direct clinical verification.

Methodologically, the lack of consistent test definitions undermines comparability across studies and complicates the interpretation of reported diagnostic accuracy. More broadly, the present findings illustrate a structural vulnerability in contemporary medical knowledge dissemination, in which imprecise descriptions may become normalized through repetition across academic, digital, and AI-mediated channels. Addressing this challenge will require more explicit test definitions, critical engagement with original sources, and improved mechanisms for standardizing clinical examination techniques. In the longer term, the development of internationally accessible reference frameworks for standardized clinical tests may help to enhance reproducibility, comparability, and patient safety in both clinical practice and research.

5. Conclusions

The present study demonstrates that inconsistent and inaccurate descriptions of the Hornblower sign are common across scientific literature, digital media, and AI-based language models. Quantitative evaluation revealed that even primary studies referencing the original description frequently conflate the Hornblower sign with the Patte test. These inaccuracies are further propagated through digital dissemination, with publicly accessible

digital sources exhibiting the lowest overall accuracy. Such inconsistency may undermine dependable clinical decision-making and limit scientific comparability.

Large language models were found to mirror prevailing descriptions rather than resolve existing inconsistencies, suggesting that inaccuracies in clinical test descriptions may be rooted in peer-reviewed literature and are subsequently reinforced by digital information systems. More broadly, these findings point to a structural challenge in contemporary medical knowledge transfer, whereby imprecise definitions become normalized through repeated citation and widespread dissemination. Once embedded in AI-based models, such imprecise definitions are propagated through model outputs, underscoring the responsibility of the scientific community to provide clear and rigorous source material.

Given the observed heterogeneity, imprecise descriptions of the Hornblower sign may compromise diagnostic interpretation and limit comparability across studies. Clear operational definitions and closer adherence to original test descriptions therefore appear essential to ensure diagnostic reliability and, in turn, patient safety. Future research should prioritize the development of internationally accessible reference frameworks for standardized clinical examination procedures to improve reproducibility and comparability.

Author Contributions: Conceptualization, P.M.; methodology, P.M. and S.Y.; validation, P.M. and S.Y.; investigation, P.M., S.Y. and S.S.; resources, P.M.; data curation, P.M. and S.Y.; writing—original draft preparation, P.M., S.Y. and S.S.; writing—review and editing, P.M., S.Y., S.S. and R.H.; visualization, S.Y.; supervision, P.M. and R.H.; project administration, P.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study did not involve human participants and therefore did not require approval from an Institutional Review Board or ethics committee.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are not publicly available due to privacy and data protection restrictions but are available from the corresponding author upon reasonable request.

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

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
LLM	Large Language Model
MRI	Magnetic Resonance Imaging

References

1. Davies, K. The information-seeking behaviour of doctors: A review of the evidence. *Health Info Libr. J.* **2007**, *24*, 78–94. [CrossRef] [PubMed]
2. Aakre, C.A.; Pencille, L.J.; Sorensen, K.J.; Shellum, J.L.; Del Fiol, G.; Maggio, L.A.; Prokop, L.J.; Cook, D.A. Electronic Knowledge Resources and Point-of-Care Learning: A Scoping Review. *Acad. Med.* **2018**, *93*, S60–S67. [CrossRef] [PubMed]
3. Youssef, Y.; Youssef, S.; Melcher, P.; Henkelmann, R.; Osterhoff, G.; Theopold, J. How Accurately Can ChatGPT 3.5 Answer Frequently Asked Questions by Patients on Glenohumeral Osteoarthritis? Obere Extremität. 15 November 2024. Available online: https://link.springer.com/10.1007/s11678-024-00836-1 (accessed on 4 March 2025).
4. Grote, T.; Berens, P. On the ethics of algorithmic decision-making in healthcare. *J. Med. Ethics* **2020**, *46*, 205–211. [CrossRef] [PubMed]
5. Baeva, A.V. Epistemic status of artificial intelligence in medical practice: Ethical challenges. *Digit. Diagn.* **2024**, *5*, 120–132. [CrossRef]

6. Bossuyt, P.M.; Reitsma, J.B.; Bruns, D.E.; Gatsonis, C.A.; Glasziou, P.P.; Irwig, L.; Lijmer, J.G.; Moher, D.; Rennie, D.; de Vet, H.C.W.; et al. STARD2015: An updated list of essential items for reporting diagnostic accuracy studies. *BMJ* **2015**, *277*, 826–832.
7. Ochodo, E.A.; De Haan, M.C.; Reitsma, J.B.; Hooft, L.; Bossuyt, P.M.; Leeflang, M.M.G. Overinterpretation and Misreporting of Diagnostic Accuracy Studies: Evidence of "Spin". *Radiology* **2013**, *267*, 581–588. [CrossRef]
8. Nissen, S.B.; Magidson, T.; Gross, K.; Bergstrom, C.T. Publication bias and the canonization of false facts. *eLife* **2016**, *5*, e21451. [CrossRef]
9. Henderson, E.L.; Simons, D.J.; Barr, D.J. The Trajectory of Truth: A Longitudinal Study of the Illusory Truth Effect. *J. Cogn.* **2021**, *4*, 29. [CrossRef]
10. Ly, D.P.; Bernstein, D.M.; Newman, E.J. An ongoing secondary task can reduce the illusory truth effect. *Front. Psychol.* **2024**, *14*, 1215432. [CrossRef]
11. Arthuis, M. Obstetrical paralysis of the brachial plexus. I. Diagnosis. Clinical study of the initial period. *Rev. Chir. Orthop. Reparatrice Appar. Mot.* **1972**, *58*, 124–126.
12. Gerber, C.; Wirth, S.H.; Farshad, M. Treatment options for massive rotator cuff tears. *J. Shoulder Elb. Surg.* **2011**, *20*, S20–S29. [CrossRef]
13. Williams, M.D.; Edwards, T.B.; Walch, G. Understanding the Importance of the Teres Minor for Shoulder Function: Functional Anatomy and Pathology. *J. Am. Acad. Orthop. Surg.* **2018**, *26*, 150–161. [CrossRef] [PubMed]
14. Patte, D.; Goutallier, D. Extensive anterior release in the painful shoulder caused by anterior impingement. *Rev. Chir. Orthop. Reparatrice Appar. Mot.* **1988**, *74*, 306–311. [PubMed]
15. Whittaker, R.L.; Alenabi, T.; Kim, S.Y.; Dickerson, C.R. Regional Electromyography of the Infraspinatus and Supraspinatus Muscles During Standing Isometric External Rotation Exercises. *Sports Health* **2022**, *14*, 725–732. [CrossRef] [PubMed]
16. Ackland, D.C.; Pandy, M.G. Moment arms of the shoulder muscles during axial rotation. *J. Orthop. Res.* **2011**, *29*, 658–667. [CrossRef]
17. Homeier, D.; Adams, M.; Lynch, T.; Cognetti, D. Inaccurate Citations Are Prevalent Within Orthopaedic Sports Medicine Literature. *Arthrosc. Sports Med. Rehabil.* **2024**, *6*, 100873. [CrossRef]
18. Gazendam, A.; Cohen, D.; Morgan, S.; Ekhtiari, S.; Ghert, M. Quotation Errors in High-Impact-Factor Orthopaedic and Sports Medicine Journals. *JBJS Open Access* **2021**, *6*, e21.00019. [CrossRef]
19. Walch, G.; Boulahia, A.; Calderone, S.; Robinson, A.H.N. The 'dropping' and 'hornblower's' signs in evaluation of rotator-cuff tears. *J. Bone Jt. Surg. Br.* **1998**, *80*, 624–628. [CrossRef]
20. Reinold, M.M.; Escamilla, R.; Wilk, K.E. Current Concepts in the Scientific and Clinical Rationale Behind Exercises for Gleno-humeral and Scapulothoracic Musculature. *J. Orthop. Sports Phys. Ther.* **2009**, *39*, 105–117. [CrossRef]
21. Youssef, S.; Brehmer, A.; Melcher, P.; Henkelmann, R.; Hepp, P.; Theopold, J.; Elze, M.; Youssef, Y.; Osterhoff, G. How accurately do large language models answer patient questions on anterior cruciate ligament tears? A comparative study. *Knee* **2025**, *58*, 104276. [CrossRef]
22. Maslej, N.; Fattorini, L.; Perrault, R.; Gil, Y.; Parli, V.; Kariuki, N.; Capstick, E.; Reuel, A.; Brynjolfsson, E.; Etchemendy, J.; et al. Artificial Intelligence Index Report 2025. *arXiv* **2025**, arXiv:2504.07139. [CrossRef]
23. Villarreal-Espinosa, J.B.; Berreta, R.S.; Allende, F.; Garcia, J.R.; Ayala, S.; Familiari, F.; Chahla, J. Accuracy assessment of ChatGPT responses to frequently asked questions regarding anterior cruciate ligament surgery. *Knee* **2024**, *51*, 84–92. [CrossRef] [PubMed]
24. Bakhsh, W.; Nicandri, G. Anatomy and Physical Examination of the Shoulder. *Sports Med. Arthrosc. Rev.* **2018**, *26*, e10–22. [CrossRef] [PubMed]
25. Panayiotou Charalambous, C. Clinical Examination of the Shoulder. In *The Shoulder Made Easy*; Springer International Publishing: Cham, Switzerland, 2019; pp. 77–122. Available online: http://link.springer.com/10.1007/978-3-319-98908-2_5 (accessed on 14 December 2025).
26. Codsi, M.; McCarron, J.; Hinchey, J.W.; Brems, J.J. Clinical evaluation of shoulder problems. In *Rockwood and Matsen's The Shoulder*, 6th ed.; Elsevier: Philadelphia, PA, USA, 2022.
27. Kim, H.M.; Dahiya, N.; Teefey, S.A.; Keener, J.D.; Yamaguchi, K. Sonography of the Teres Minor: A Study of Cadavers. *Am. J. Roentgenol.* **2008**, *190*, 589–594. [CrossRef]
28. Bi, A.S.; Jazrawi, L.M.; Cohen, S.B.; Erickson, B.J. The Physical Examination of the Throwing Shoulder. *Clin. Sports Med.* **2025**, *44*, 113–128. [CrossRef]
29. Razmjou, H. Cuff Tear Arthropathy. In *Clinical and Radiological Examination of the Shoulder Joint*; Springer International Publishing: Cham, Switzerland, 2022; pp. 59–74. Available online: https://link.springer.com/10.1007/978-3-031-10470-1_5 (accessed on 14 December 2025).
30. Razmjou, H. Diagnostic Accuracy of Clinical Tests in Detecting Rotator Cuff Pathology. *Orthop. Sports Med. Open Access J.* **2019**, *2*, 170–179. [CrossRef]
31. Kim, K.C.; Rhee, K.J.; Shin, H.D.; Byun, K.Y. Physical Examinations of Rotator Cuff Tear. *J. Korean Shoulder Elb. Soc.* **2008**, *11*, 13–18. [CrossRef]

32. Pandey, V. Diagnostic Clinical Tests in Rotator Cuff Tear: Which and Why? *Asian J. Arthrosc.* **2021**, *6*, 24–30. [CrossRef]
33. Robinson, S.; Oakes, N.; Shahane, S. History Taking and Clinical Assessment of the Shoulder. In *Textbook of Shoulder Surgery*; Trail, I.A., Funk, L., Rangan, A., Nixon, M., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 555–586. Available online: http://link.springer.com/10.1007/978-3-319-70099-1_34 (accessed on 14 December 2025).
34. Blønd, L. Skulderundersøgelser—et Kompendium. Herlev. Available online: https://larsblond.com/wp-content/uploads/2016/04/Skulder-kompendium-med-indholdsfortegnelse1.pdf (accessed on 14 December 2025).
35. Agarwal, A.; Arafa, M.; Avidor-Reiss, T.; Hamoda, T.A.A.A.M.; Shah, R. Citation Errors in Scientific Research and Publications: Causes, Consequences, and Remedies. *World. J. Mens Health* **2023**, *41*, 461. [CrossRef]
36. Simkin, M.V.; Roychowdhury, V.P. Read Before You Cite! *arXiv* **2002**, arXiv:cond-mat/0212043. [CrossRef]
37. The Decision Lab. Authority Bias—Why Do We Always Trust the Doctor, Even Though They Might Be Wrong? The Decision Lab. Available online: https://thedecisionlab.com/biases/authority-bias (accessed on 21 December 2025).
38. Keller, R. Grundlagen der Wissenssoziologischen Diskursanalyse. In *Diskursanalyse für die Kommunikationswissenschaft*; Wiedemann, T., Lohmeier, C., Eds.; Springer Fachmedien Wiesbaden: Wiesbaden, Germany, 2019; pp. 35–60. Available online: http://link.springer.com/10.1007/978-3-658-25186-4_3 (accessed on 25 December 2025).
39. Strzelecki, A. Google Medical Update: Why Is the Search Engine Decreasing Visibility of Health and Medical Information Websites? *Int. J. Environ. Res. Public Health* **2020**, *17*, 1160. [CrossRef]
40. Haslam, K.; Doucette, H.; Hachey, S.; MacCallum, T.; Zwicker, D.; Smith-Brilliant, M.; Gilbert, R. YouTube videos as health decision aids for the public: An integrative review. *Can. J. Dent. Hyg.* **2019**, *53*, 53–66.
41. Wong, L.; Ali, A.; Xiong, R.; Shen, S.Z.; Kim, Y.; Agrawal, M. Retrieval-augmented systems can be dangerous medical communicators. *arXiv* **2025**, arXiv:2502.14898.
42. Kurokawa, D.; Sano, H.; Nagamoto, H.; Omi, R.; Shinozaki, N.; Watanuki, S.; Kishimoto, K.N.; Yamamoto, N.; Hiraoka, K.; Tashiro, M.; et al. Muscle activity pattern of the shoulder external rotators differs in adduction and abduction: An analysis using positron emission tomography. *J. Shoulder Elb. Surg.* **2014**, *23*, 658–664. [CrossRef]
43. Ochs, K. Die Rotatorenmanschettenruptur—eine Berufserkrankung?—eine Epidemiologische Studie. Ph.D. Thesis, Julius-Maximilians-Universität Würzburg, Würzburg, Germany, 2008. Available online: https://opus.bibliothek.uni-wuerzburg.de/opus4-wuerzburg/frontdoor/deliver/index/docId/2732/file/Diss_KBO.pdf (accessed on 26 December 2025).