

Review

The Temporal Evolution of Large Language Model Performance: A Comparative Analysis of Past and Current Outputs in Scientific and Medical Research

Ishith Seth ^{1,2} , Gianluca Marcaccini ^{3,*} , Bryan Lim ¹ , Jennifer Novo ⁴ , Stephen Bacchi ⁵ ,
Roberto Cuomo ³ , Richard J. Ross ^{1,2} and Warren M. Rozen ^{1,2} 

- ¹ Department of Plastic and Reconstructive Surgery, Peninsula Health, Frankston, VIC 3199, Australia; lim.bryan58@gmail.com (B.L.); ross@phcn.vic.gov.au (R.J.R.); warrenrozen@hotmail.com (W.M.R.)
² Department of Plastic and Reconstructive Surgery, Austin Health, Heidelberg, VIC 3084, Australia
³ Department of Plastic and Reconstructive Surgery, University of Siena, 53100 Siena, Italy; robertocuomo@outlook.com
⁴ Faculty of Medicine and Surgery, The University of Notre Dame, Chippendale, NSW 2008, Australia; jenny.novo@gmail.com
⁵ Harvard Medical School, Massachusetts General Hospital, Boston, MA 02144, USA; stephenbacchi@gmail.com
* Correspondence: gianlu32@gmail.com

Abstract

Background: Large language models (LLMs) such as ChatGPT have evolved rapidly, with notable improvements in coherence, factual accuracy, and contextual relevance. However, their academic and clinical applicability remains under scrutiny. This study evaluates the temporal performance evolution of LLMs by comparing earlier model outputs (GPT-3.5 and GPT-4.0) with ChatGPT-4.5 across three domains: aesthetic surgery counseling, an academic discussion base of thumb arthritis, and a systematic literature review. **Methods:** We replicated the methodologies of three previously published studies using identical prompts in ChatGPT-4.5. Each output was assessed against its predecessor using a nine-domain Likert-based rubric measuring factual accuracy, completeness, reference quality, clarity, clinical insight, scientific reasoning, bias avoidance, utility, and interactivity. Expert reviewers in plastic and reconstructive surgery independently scored and compared model outputs across versions. **Results:** ChatGPT-4.5 outperformed earlier versions across all domains. Reference quality improved most significantly (a score increase of +4.5), followed by factual accuracy (+2.5), scientific reasoning (+2.5), and utility (+2.5). In aesthetic surgery counseling, GPT-3.5 produced generic responses lacking clinical detail, whereas ChatGPT-4.5 offered tailored, structured, and psychologically sensitive advice. In academic writing, ChatGPT-4.5 eliminated reference hallucination, correctly applied evidence hierarchies, and demonstrated advanced reasoning. In the literature review, recall remained suboptimal, but precision, citation accuracy, and contextual depth improved substantially. **Conclusion:** ChatGPT-4.5 represents a major step forward in LLM capability, particularly in generating trustworthy academic and clinical content. While not yet suitable as a standalone decision-making tool, its outputs now support research planning and early-stage manuscript preparation. Persistent limitations include information recall and interpretive flexibility. Continued validation is essential to ensure ethical, effective use in scientific workflows.

Keywords: large language models; ChatGPT; scientific writing; evidence-based medicine; artificial intelligence in healthcare



Academic Editor: Antony Bryant

Received: 5 June 2025

Revised: 8 August 2025

Accepted: 21 August 2025

Published: 26 August 2025

Citation: Seth, I.; Marcaccini, G.; Lim, B.; Novo, J.; Bacchi, S.; Cuomo, R.; Ross, R.J.; Rozen, W.M. The Temporal Evolution of Large Language Model Performance: A Comparative Analysis of Past and Current Outputs in Scientific and Medical Research. *Informatics* **2025**, *12*, 86.

<https://doi.org/10.3390/informatics12030086>

Copyright: © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license

(<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Over the past few years, the evolution of large language models (LLMs), supported by advanced machine learning techniques such as transformer-based architectures, reinforcement learning from human feedback (RLHF), large-scale unsupervised pretraining, and instruction tuning, has fundamentally transformed the way scientific research is conducted, written, and disseminated. These advancements have led to measurable improvements in coherence, factual accuracy, and contextual understanding [1,2]. Advanced machine learning techniques have propelled models such as ChatGPT from early experimental tools to sophisticated systems capable of generating complex, nuanced text. These models have been increasingly employed to support literature searches, data synthesis, and even the preliminary drafting of scientific manuscripts [3]. Early evaluations, however, revealed notable limitations, including superficial analyses, occasional factual inaccuracies, and instances of fabricated references. Such shortcomings raised important questions regarding AI-generated content's reliability and academic utility [3–5].

The primary aim of the present study is to conduct a comprehensive comparative analysis of LLM outputs by replicating and extending methodologies from three seminal studies. The first of these studies, published in May 2023 and based on outputs generated in January 2023, offered an early perspective on how LLMs performed when applied to clinically oriented research questions [3]. The second study, published in October 2024 with outputs generated in May 2023, further explored the applicability of LLMs in addressing specific medical queries [6]. The third study, a literature review with outputs generated in April 2024, examined the evolution of systematic literature search capabilities and highlighted improvements in content depth and reference accuracy [7]. In our investigation, we strictly adhere to the original experimental protocols by using the same set of prompts as employed in the previous studies. Because the original studies utilised only ChatGPT, our analysis employs the latest version, ChatGPT 4.5, to ensure a direct and meaningful comparison. Our analysis is structured around several key performance dimensions. We first assess the accuracy of responses by examining the factual correctness and their alignment with the current state of scientific knowledge. Next, we evaluate the completeness of the information provided by determining whether the responses comprehensively address all aspects of the queries. We also analyse the coherence and clarity of the generated text, focusing on logical flow and overall readability. Furthermore, we investigate the incidence of bias and errors, with particular attention to occurrences of fabricated or misleading references. Finally, we consider the practical utility of the outputs in the context of scientific and medical research, including their capacity to support literature reviews, data synthesis, and academic writing.

It is important to note that early iterations of LLMs, as reflected in studies [3,6], were characterised by a tendency toward generic or surface-level responses. These limitations were primarily attributable to constraints in training data and model architecture at that time and were empirically demonstrated in domains such as literature searching, where ChatGPT significantly underperformed against human researchers [8]. Recent advancements have addressed many of these issues. For example, ChatGPT 4.5 exhibits marked improvements in the depth and precision of its responses, offering more nuanced and contextually relevant information than its predecessors [3,6]. The implications of this study are significant for integrating AI technologies in academic research. By systematically comparing historical outputs with those generated by the most advanced models available today, we aim to depict progress made over the past 12 to 24 months. Our findings are expected to illuminate both the benefits and the persistent challenges associated with using LLMs, informing future research directions and guiding best practices for ethical implementation [1,5]. Ultimately, this work seeks to bolster confidence in using AI-generated content

while delineating the boundaries within which these technologies can most effectively support scientific inquiry.

2. Materials and Methods

This study was designed to evaluate the evolution in the performance of LLMs over time by replicating and extending the methodologies of the author's three previously published studies. The research adopts a comparative, qualitative design, focusing on outputs generated by earlier versions of ChatGPT (GPT-3.5 and GPT-4.5) and comparing them to outputs produced by the most recent iterations, ChatGPT4o. The objective was to assess improvements in accuracy, completeness, reference reliability, clarity, and overall practical utility in scientific and medical research.

To ensure methodological rigour and consistency, we selected three benchmark studies in which earlier LLMs had been tasked with specific academic or clinical prompts. These included: (1) a rhinoplasty consultation simulation involving nine standardised patient-focused questions derived from the American Society of Plastic Surgeons' checklist; (2) an academic conversation evaluating the use of implants in the management of base of thumb arthritis, structured as five iterative scientific prompts; and (3) a systematic literature search comparing ChatGPT and other AI platforms against human researcher performance in identifying high-level evidence related to trapeziometacarpal joint osteoarthritis.

In the present study, we recreated each original scenario using ChatGPT-4.5, replicating the prompts used in the prior investigations without modification. These prompts reflected the nature of the original tasks: (1) a structured set of nine patient-focused questions for aesthetic surgery counseling; (2) five sequential, evidence-based academic questions on the surgical management of base of thumb arthritis; and (3) a systematic literature search query formulated with predefined inclusion criteria and Boolean logic elements. All prompts were entered verbatim in a single ChatGPT-4.5 session, in English, without rephrasing or multiple response generation, to ensure consistency with the original studies. All prompts were entered into a single ChatGPT-4.5 session under consistent conditions to eliminate variation due to model sampling or user behavior. The model was instructed not to generate multiple versions of each response. Responses were fully collected and preserved in their original form for blinded assessment.

Each ChatGPT-4.5 output was evaluated using a predefined rubric based on five performance domains: (1) factual accuracy, defined as the correctness of scientific or clinical information provided; (2) completeness, reflecting the model's ability to address all aspects of the prompt; (3) reference quality, including the presence or absence of hallucinated or unverifiable citations; (4) clarity and coherence, judged by logical flow, grammatical accuracy, and readability; and (5) practical utility, particularly in the context of research drafting, literature review, or patient counseling. The prompts used in this study were identical to those reported in full in our previously published studies [3,6,8] and are also displayed in the comparative tables within this manuscript (Tables 1 and 2). No additional modifications or variations were introduced. All ChatGPT-4.5 outputs were generated between March 2025 in a single session per domain, ensuring consistent model behaviour during testing.

Table 1. Aesthetic Surgery Advice and Counseling from Artificial Intelligence: A Rhinoplasty Consultation with ChatGPT.

(a)		
Prompt	Response by ChatGPT-3	Response by ChatGPT-4.5
"I'm thinking of getting a rhinoplasty, am I a good candidate for this procedure?"	"As an AI language model, I cannot make individual medical assessments or provide medical advice. However, generally, a good candidate for rhinoplasty is someone who is in good overall health, has a fully developed nasal bone structure, and has realistic expectations about the outcome of the surgery. It is important to have a consultation with a qualified plastic surgeon to discuss your specific case and determine if rhinoplasty is right for you. The surgeon will consider factors such as the shape of your nose, your facial features, and your overall health to determine if you are a good candidate for the procedure."	"Great question! Whether you're a good candidate for rhinoplasty depends on a few key factors. Here's what surgeons usually look for: 1. Overall Health: You should be in good physical health, without serious medical conditions that could impair healing. 2. Age: Ideally, your facial growth should be complete. This typically means age 16 or older, depending on individual development. 3. Realistic Expectations: It's important to have clear, realistic goals and understand what the procedure can and cannot achieve. 4. Non-Smoker: Smoking can impair healing, so non-smokers—or those willing to quit before and after surgery—are better candidates. 5. Functional Concerns: If you have breathing difficulties due to structural issues (like a deviated septum), rhinoplasty might also be medically beneficial. 6. Emotional Readiness: It's important to pursue the surgery for personal reasons, not due to external pressure. Would you like help figuring out how to prepare for a consultation or what questions to ask a surgeon?"
"How might the surgeon perform my procedure?"	"The specific details of how a rhinoplasty procedure is performed will depend on your individual case and your surgeon's preferences. However, in general, there are two techniques that may be used for rhinoplasty: open rhinoplasty and closed rhinoplasty." Open rhinoplasty involves making a small incision on the outside of the nose, along the strip of tissue that separates the nostrils (called the columella), and then lifting the skin to access the underlying nasal structures. This technique provides the surgeon with a more direct view of the nasal structures and may be preferred for more complex procedures. Closed rhinoplasty involves making incisions inside the nostrils, so that all of the surgical work is performed through these incisions. This technique may be preferred for less complex procedures and may result in less visible scarring. Your surgeon will evaluate your individual case and discuss with you which technique they recommend for your rhinoplasty procedure. They will also discuss the specifics of the surgery, including how long it will take, what type of anesthesia will be used, and what you can expect during and after the surgery."	"The surgical approach for a rhinoplasty depends on your goals (cosmetic, functional, or both) and your nasal anatomy. Generally, surgeons use one of two main techniques: 1. Closed Rhinoplasty (Endonasal) · Incisions are made inside the nostrils, so there are no visible external scars. · Used for minor reshaping, tip refinement, or dorsal hump reduction. · Advantages: Less swelling, no visible scar, potentially shorter recovery. 2. Open Rhinoplasty · Incisions are made inside the nostrils plus a small one across the columella (the strip of tissue between the nostrils). · Used for more complex cases (major reshaping, asymmetry, revision surgery). · Advantages: Better visibility and precision for the surgeon, especially for tip work. Common Surgical Steps: · Dorsal reduction or augmentation (removing or adding cartilage/bone to smooth the nasal bridge). · Tip reshaping (refining the angle or projection of the nasal tip). · Septoplasty (if needed to straighten the septum for better breathing). · Osteotomies (controlled bone fractures to narrow or straighten the nose). The surgeon may use your own cartilage (usually from the septum, ear, or rib) for grafting if needed. Would you like an example of how a surgeon might plan the surgery based on a specific nasal feature (e.g., hump, wide tip, breathing issues)?"

Table 1. Cont.

(a)		
Prompt	Response by ChatGPT-3	Response by ChatGPT-4.5
“What are the steps in a open rhinoplasty? Please be specific”	“In an open rhinoplasty, the following steps are typically involved: Anesthesia: The patient is given general anesthesia, which ensures that they are comfortable and asleep during the procedure. Incision: A small incision is made on the outside of the nose, along the strip of tissue that separates the nostrils (called the columella). Dissection: The skin is lifted away from the underlying nasal structures, providing the surgeon with a clear view of the nasal anatomy.” Reshaping the nasal structures: Using various surgical techniques, the surgeon reshapes the nasal structures to achieve the desired result. This may involve removing or adding tissue, reshaping the cartilage or bone, or adjusting the position of the nasal septum. Closing the incision: Once the necessary adjustments have been made, the skin is re-draped over the underlying structures, and the incision is closed with sutures. Application of a splint: A splint or cast is applied to the nose to help it maintain its new shape and protect it during the initial healing period. After the procedure, the patient will be monitored in a recovery area until they are alert and comfortable. They may experience swelling, bruising, and discomfort for several days to several weeks following the surgery, and they will need to follow their surgeon’s postoperative instructions carefully to ensure proper healing. The surgeon will provide the patient with specific instructions on caring for their nose during the recovery period and when to follow up for a post-operative visit.”	“Absolutely—here’s a step-by-step breakdown of an open rhinoplasty, described in detail so you understand exactly what the procedure entails: 1. Anesthesia · The procedure begins under general anesthesia or local anesthesia with sedation, depending on the surgeon’s preference and the complexity of the case. 2. Incision · The surgeon makes: · Two intra-nasal incisions inside each nostril. · One trans-columellar incision across the columella (the strip of skin between the nostrils), connecting the internal cuts. · This forms a “stair-step” or “inverted V” shape to allow the skin to be lifted. 3. Skin Elevation · The nasal skin and soft tissue are carefully lifted off the underlying cartilage and bone, exposing the nasal framework. · This gives the surgeon full visibility of the nasal anatomy, especially the tip cartilages (lower lateral cartilages) and nasal dorsum. 4. Structural Modification This is the core of the surgery and varies based on goals: a. Dorsal Hump Reduction · If needed, the bony-cartilaginous hump is shaved down using: · A scalpel or rasp for cartilage, · Osteotomes or burrs for bone. b. Tip Reshaping · The lower lateral cartilages (tip cartilages) are modified: · Can be resected, sutured, or repositioned. · Sometimes grafts (e.g., columellar strut, shield graft) are used for better support or shape. c. Nasal Width Adjustment · Osteotomies (precise fractures of the nasal bones) may be done to narrow or straighten the nasal bridge. d. Septoplasty (if needed) · If the septum is deviated, the surgeon will straighten it and may harvest cartilage grafts from it. e. Grafting · Cartilage grafts may be placed to support the tip, improve symmetry, or reinforce weakened areas. · Sources: septum (most common), ear, or rib (in complex or revision cases). 5. Closure · The skin is redraped over the modified framework. · The columellar incision is closed with fine sutures (usually non-dissolvable for external ones, dissolvable for internal). · Care is taken to ensure alignment and symmetry. 6. Dressing & Splinting · Nasal splints or packing may be placed inside the nose (especially if septoplasty was done). · An external nasal splint (thermoplastic or aluminum) is applied over the bridge to protect the structure and reduce swelling. Would you like a visual diagram or a specific version of this tailored to a particular nasal concern (like a droopy tip or bulbous nose)?”

Table 1. Cont.

(a)														
Prompt	Response by ChatGPT-3	Response by ChatGPT-4.5												
“What are the steps in an closed rhinoplasty? Please be specific”	“In a closed rhinoplasty, the following steps are typically involved: Anesthesia: The patient is given general anesthesia, which ensures that they are comfortable and asleep during the procedure. Incisions: Incisions are made inside the nostrils, and the surgeon accesses the nasal structures through these incisions. Dissection: The skin is lifted away from the underlying nasal structures, providing the surgeon with a clear view of the nasal anatomy. Reshaping the nasal structures: Using various surgical techniques, the surgeon reshapes the nasal structures to achieve the desired result. This may involve removing or adding tissue, reshaping the cartilage or bone, or adjusting the position of the nasal septum. Closing the incisions: Once the necessary adjustments have been made, the incisions inside the nostrils are closed with sutures.”	“Certainly! A closed rhinoplasty (also called endonasal rhinoplasty) is performed entirely through incisions made inside the nostrils, with no external scar. It requires more technical precision due to the limited visibility, and it’s typically used for more subtle reshaping. Here’s a detailed step-by-step breakdown of a closed rhinoplasty: 1. Anesthesia · Performed under general anesthesia or local anesthesia with sedation, depending on case complexity and patient/surgeon preference. 2. Internal Incisions · The surgeon makes intra-nasal incisions: · Along the marginal incision (inside the nostril rim). · Possibly extending to an intercartilaginous incision (between the upper and lower lateral cartilages). · These allow access to the nasal tip, dorsum, and septum without lifting the nasal skin externally. 3. Limited Skin Elevation · Through the internal access points, the surgeon carefully elevates soft tissue over the nasal structures using fine instruments. · The skin envelope stays mostly intact, reducing trauma and postoperative swelling. 4. Structural Modification All structural work is done through the small internal openings: a. Dorsal Hump Reduction · Cartilage and bone forming the dorsal hump are shaved or rasped. · This is done by inserting instruments through the nostrils and working beneath the skin. b. Tip Refinement · The lower lateral cartilages (tip cartilages) are accessed indirectly. · Tip reshaping may involve: · Cephalic trim (removing part of the cartilage), · Suturing techniques (like interdomal or transdomal sutures), · Minimal cartilage grafts (if accessible through the nostrils). c. Nasal Bone Narrowing · Osteotomies (fractures of the nasal bones) may still be done through internal incisions using a guarded chisel to narrow or straighten the bridge. d. Septoplasty (if needed) · A submucosal resection may be performed to straighten the septum and improve airflow. · The harvested septal cartilage can also be used for grafts. 5. Closure · All internal incisions are closed with dissolvable sutures. · Since no external incisions are made, no visible scars are left. 6. Splinting & Dressing · An external nasal splint is typically applied for 5–7 days to stabilize the bones and reduce swelling. · Internal splints or packing may be placed if septal work was done. Summary of Key Features: <table><tr><td>Feature</td><td>Closed Rhinoplasty</td></tr><tr><td>Incision location</td><td>Inside nostrils only</td></tr><tr><td>External scar</td><td>None</td></tr><tr><td>Visibility for surgeon</td><td>Limited</td></tr><tr><td>Best for</td><td>Minor/moderate reshaping</td></tr><tr><td>Recovery</td><td>Often quicker, less swelling</td></tr></table>	Feature	Closed Rhinoplasty	Incision location	Inside nostrils only	External scar	None	Visibility for surgeon	Limited	Best for	Minor/moderate reshaping	Recovery	Often quicker, less swelling
	Feature	Closed Rhinoplasty												
Incision location	Inside nostrils only													
External scar	None													
Visibility for surgeon	Limited													
Best for	Minor/moderate reshaping													
Recovery	Often quicker, less swelling													
Let me know if you’d like a comparison with open rhinoplasty or if you want this tailored to a specific nasal issue (like crooked nose or wide tip).”														

Table 1. Cont.

(a)																
Prompt	Response by ChatGPT-3	Response by ChatGPT-4.5														
“Under what circumstances would you perform one or the other?”	“The decision to perform an open or closed rhinoplasty will depend on the individual case and the specific goals of the surgery. Open rhinoplasty is typically used for more complex cases or when more precision is needed. It provides the surgeon with a more direct view of the nasal structures, allowing for greater accuracy in making adjustments. Open rhinoplasty may also be used when significant changes need to be made to the tip of the nose or when major reconstruction is required. Closed rhinoplasty may be used for less complex cases, such as when minor adjustments are needed to the nasal tip or bridge. Closed rhinoplasty may also be preferred when the patient is concerned about visible scarring, as the incisions are made inside the nostrils and are therefore not visible externally. Ultimately, the decision to perform an open or closed rhinoplasty will depend on a variety of factors, including the patient’s individual needs and goals, the surgeon’s expertise and preference, and the complexity of the surgery. The surgeon will evaluate the patient’s individual case and recommend the technique that is best suited to achieve the desired result.”	“Great question. The choice between open and closed rhinoplasty depends on multiple factors, including the complexity of the case, the surgeon’s expertise, and the specific goals of the patient. Here’s a breakdown of when a surgeon might choose one over the other: Closed Rhinoplasty (Endonasal) Preferred when: · The corrections are mild to moderate, such as: · Small dorsal hump reduction · Minor tip refinement · Narrowing the nasal bones · There’s no need for major cartilage grafting or structural work · The patient desires a shorter recovery and no external scar · It’s a primary (first-time) rhinoplasty with no complex anatomical issues · Tip asymmetry is minimal Advantages: · No visible scar · Possibly faster recovery · Less swelling post-op Limitations: · Limited visibility of nasal structures · More technically demanding for major changes · Harder to perform precise tip work or graft placement Open Rhinoplasty Preferred when: · The case is complex or revisionary, such as: · Significant tip reshaping or projection changes · Asymmetry or deformities (e.g., cleft nose, trauma) · Need for multiple grafts or major structural reconstruction · Complex deviated septum or crooked nose · Precise control and visibility are essential (e.g., for revision rhinoplasty) Advantages: · Full exposure of nasal anatomy · Allows precise manipulation of cartilage and bones · Better for teaching/training situations Limitations: · Small scar on the columella (usually fades well) · Slightly longer recovery, especially due to swelling at the nasal tip In Summary: <table><tr><td>Clinical Scenario</td><td>Likely Approach</td></tr><tr><td>Minor hump + narrow tip</td><td>Closed rhinoplasty</td></tr><tr><td>Complex tip work (e.g., droopy or bulbous)</td><td>Open rhinoplasty</td></tr><tr><td>Revision surgery</td><td>Open rhinoplasty</td></tr><tr><td>Functional + cosmetic goals</td><td>Either, depends on anatomy</td></tr><tr><td>Trauma or asymmetry</td><td>Often open rhinoplasty</td></tr><tr><td>Patient prioritizes no external scar</td><td>Closed (if feasible)</td></tr></table>	Clinical Scenario	Likely Approach	Minor hump + narrow tip	Closed rhinoplasty	Complex tip work (e.g., droopy or bulbous)	Open rhinoplasty	Revision surgery	Open rhinoplasty	Functional + cosmetic goals	Either, depends on anatomy	Trauma or asymmetry	Often open rhinoplasty	Patient prioritizes no external scar	Closed (if feasible)
	Clinical Scenario	Likely Approach														
	Minor hump + narrow tip	Closed rhinoplasty														
	Complex tip work (e.g., droopy or bulbous)	Open rhinoplasty														
	Revision surgery	Open rhinoplasty														
Functional + cosmetic goals	Either, depends on anatomy															
Trauma or asymmetry	Often open rhinoplasty															
Patient prioritizes no external scar	Closed (if feasible)															
	If you have a particular nasal feature you’re curious about (like a wide tip, drooping tip, breathing difficulty, etc.), I can explain which approach fits best and why.”															

Table 1. Cont.

(a)		
Prompt	Response by ChatGPT-3	Response by ChatGPT-4.5
“What are the risks and complications associated with this procedure, and how are they handled?”	“Like any surgical procedure, rhinoplasty has risks and potential complications. Some risks include: · Bleeding Infection · Adverse reactions to anesthesia · Scarring · Numbness or tingling · Difficulty breathing · Unsatisfactory aesthetic results · Need for revision surgery Your surgeon will discuss these risks with you in detail during your consultation, as well as any other risks that may be specific to your individual case. To minimize the risk of complications, it is important to choose a qualified and experienced plastic surgeon who has a thorough understanding of rhinoplasty techniques and who can provide you with realistic expectations about the outcome of the surgery. If complications do occur, your surgeon will work to address them promptly and effectively. Depending on the nature of the complication, this may involve additional medical treatment or revision surgery. It is important to follow all of your surgeon’s instructions before and after the procedure to minimize the risk of complications and to ensure the best possible outcome.”	“Rhinoplasty, like any surgery, carries risks and potential complications. While it’s generally safe when performed by an experienced surgeon, it’s important to be informed. Here’s a detailed list of the most common and serious risks, along with how they are usually handled: 1. Swelling and Bruising · What it is: Common post-op effects, especially around the eyes and nose. · Management: Elevation of the head · Cold compresses in the first 48 h · Usually resolves within 2–4 weeks (tip swelling can last longer) 2. Bleeding (Epistaxis) · What it is: Mild bleeding is common in the first few days. · Management: Nasal packing (temporary) · Avoidance of aspirin/NSAIDs pre-op · Rarely, cauterization or re-packing is needed 3. Infection · What it is: Rare in rhinoplasty due to good blood supply. · Management: Prophylactic antibiotics · Local wound care if needed · Drainage and targeted antibiotics if abscess forms 4. Breathing Difficulties · What it is: Can occur if internal support structures are weakened. · Management: Septoplasty or spreader grafts during surgery to preserve airway · In rare cases, revision surgery may be needed 5. Asymmetry or Unsatisfactory Aesthetic Outcome · What it is: Nose may look uneven or not meet expectations. · Management: Small irregularities may improve over time or be treated with steroid injections or filler · Significant issues may require revision rhinoplasty (usually after 6–12 months) 6. Numbness or Sensory Changes · What it is: Temporary numbness, especially at the tip of the nose. · Management: Usually resolves in weeks to months · Rarely permanent 7. Scarring (in open rhinoplasty) · What it is: Small scar on the columella · Management: Usually fades well over time · Silicone gel, massage, or laser if scar is prominent 8. Septal Perforation · What it is: A hole in the septum, which can cause whistling or crusting. · Management: Prevented by careful technique and preserving mucosa · Small ones may heal or be managed with saline sprays · Larger ones might need surgical repair 9. Skin Issues · What it is: Skin necrosis (rare) or acne/irritation from the nasal splint · Management: Careful splint application · Monitoring and wound care if necrosis occurs 10. Anesthesia Risks · What it is: Reaction to sedation or general anesthesia · Management: Pre-operative screening · Experienced anesthesiology team Would you like a version of this tailored to your concerns or aesthetic goals? Or maybe a checklist of questions to ask the surgeon about risk management?”

Table 1. Cont.

(a)		
Prompt	Response by ChatGPT-3	Response by ChatGPT-4.5
“How long of a recovery period can I expect, and what kind of help will I need during my recovery?”	“The recovery period for rhinoplasty can vary depending on the extent of the surgery and your individual healing process. In general, you can expect to experience some swelling, bruising, and discomfort for several days to several weeks following the procedure. During the first few days following the procedure, you may need to wear a splint or cast on your nose to protect it and help it maintain its new shape. You may also experience some bleeding, nasal congestion, and difficulty breathing, but these symptoms should gradually improve over time.	“Recovery from rhinoplasty happens in stages, and while you’ll likely be up and walking the same day, full recovery takes several months. Here’s a detailed timeline of what to expect and the kind of help you might need: Immediate Post-Op (Days 1–7) What to expect: · Swelling and bruising, especially around the eyes · Nasal congestion, mild bleeding, or drainage · You’ll have an external splint and possibly internal packing or splints Help you may need: · Someone to drive you home and stay with you the first night · Help preparing meals, taking medications, and keeping your head elevated · Avoiding bending, lifting, or blowing your nose Early Recovery (Week 1–2) What to expect: · Splint and any external sutures removed at ~Day 5–7 · Bruising and swelling begin to subside · You can usually return to light, non-strenuous work after 7–10 days Help you may need: · Gentle reminders to avoid touching your nose · Possibly help with errands or childcare if you feel fatigued Intermediate Recovery (Weeks 3–6) What to expect: · Swelling continues to go down, but the tip may still feel firm or swollen · Most people feel confident returning to social activities after 3–4 weeks · Avoid strenuous exercise, contact sports, and glasses resting on your nose Help you may need: · Minimal; most people are independent again, just need patience for swelling Long-Term Recovery (Months 2–12) What to expect: · The final shape of your nose slowly reveals itself · Swelling of the nasal tip is the last to resolve (can take 6–12 months) · You’ll continue to have follow-up visits to monitor healing Help you may need: · None typically, though emotional support is valuable if you’re adjusting to changes Tips for Smoother Recovery: · Sleep with your head elevated for 1–2 weeks · Avoid blowing your nose for at least 2–3 weeks · Use saline sprays if prescribed · Follow all surgeon instructions closely, especially regarding activity and skincare Would you like a printable recovery checklist or a timeline with reminders by week?”
	Your surgeon may recommend that you take time off from work or other activities for a period of time following the surgery to allow for proper healing. You may also need to avoid certain activities, such as heavy lifting or exercise, for a period of time. It is important to follow your surgeon’s post-operative instructions closely to ensure the best possible outcome and to minimize the risk of complications. You may also want to have someone available to help you during your initial recovery period, as you may need assistance with daily tasks or transportation. Your surgeon will provide you with specific instructions on how to care for your nose during the recovery period and what to expect during each stage of the healing process. They will also schedule follow-up appointments to monitor your progress and ensure that you are healing properly.”	

Table 1. Cont.

(a)		
Prompt	Response by ChatGPT-3	Response by ChatGPT-4.5
“How can I expect my nose to look over time?”	“The recovery period for rhinoplasty can vary depending on the extent of the surgery and your individual healing process. In general, you can expect to experience some swelling, bruising, and discomfort for several days to several weeks following the procedure. During the first few days following the procedure, you may need to wear a splint or cast on your nose to protect it and help it maintain its new shape. You may also experience some bleeding, nasal congestion, and difficulty breathing, but these symptoms should gradually improve over time. Your surgeon may recommend that you take time off from work or other activities for a period of time following the surgery to allow for proper healing. You may also need to avoid certain activities, such as heavy lifting or exercise, for a period of time. It is important to follow your surgeon’s post-operative instructions closely to ensure the best possible outcome and to minimize the risk of complications. You may also want to have someone available to help you during your initial recovery period, as you may need assistance with daily tasks or transportation. Your surgeon will provide you with specific instructions on how to care for your nose during the recovery period and what to expect during each stage of the healing process. They will also schedule follow-up appointments to monitor your progress and ensure that you are healing properly.”	“Great question—rhinoplasty results evolve significantly over time, and understanding the timeline helps set realistic expectations. Here’s a general timeline of how your nose will look after surgery: Immediately After Surgery (Day 1–7) • Appearance: You’ll have a splint on your nose, possibly bruising and swelling around the eyes and cheeks. • The nose may look larger or distorted due to swelling and dressings. • You won’t be able to see the real results yet. After Splint Removal (Week 1–2) • Appearance: Initial swelling begins to go down. • The general shape of your new nose starts to appear, especially from the side. • However, the nasal tip and bridge may still look puffy or “boxy.” Early Healing (Weeks 3–6) • Appearance: Bruising is mostly gone; mild swelling remains. • In photos or from a distance, your nose might already look “done.” • Up-close, the tip can still appear bulky or firm—this is normal. Intermediate Stage (Months 2–6) • Appearance: The swelling continues to subside gradually. • The nasal bridge refines first; tip definition improves slowly. • As tissues settle, the nose looks more natural and integrated with your face. Final Result (Months 9–12+) • Appearance: The final shape of your nose becomes visible. • The tip is more defined, swelling is nearly or fully gone. • If you had thick skin or major reshaping, full refinement might take up to 18 months. Additional Notes: • Morning swelling is common in the first few months and fades throughout the day. • If cartilage grafts were used (e.g., for the tip or bridge), they’ll become more noticeable as swelling decreases. • Minor asymmetries or irregularities can become visible over time—most are subtle and normal. Would you like a set of example photos that show typical progression, or suggestions on how to track your own healing through photos?”
	“If you are dissatisfied with the cosmetic outcome of your nose surgery, you may be able to discuss revision surgery with your surgeon. Revision surgery is a secondary procedure that is performed to correct issues with the initial surgery, such as asymmetry or unsatisfactory aesthetic results. However, it is important to have realistic expectations about the outcome of revision surgery, as it may not be able to fully correct all issues. Revision surgery can be more complex than the initial surgery, and the recovery period may be longer. Before considering revision surgery, it is important to discuss your concerns with your surgeon and to give them the opportunity to address any issues with the initial surgery. They may be able to suggest alternative treatments or adjustments that can improve your satisfaction with the outcome. It is also important to choose a qualified and experienced plastic surgeon for both the initial surgery and any revision surgery. They will be able to provide you with a realistic understanding of what to expect and to help you achieve the best possible outcome.”	“If you’re unhappy with the cosmetic outcome of your rhinoplasty, you’re definitely not alone—many patients seek revision for a variety of reasons. Here’s a breakdown of your options and what to consider: 1. Give It Time • First step: Wait for the healing process to complete—especially if the issue is mild. • It can take 12–18 months for all swelling (especially in the tip) to resolve and the final shape to appear. • What looks like asymmetry or fullness early on often improves naturally. 2. Non-Surgical Options (If Minor) • If the dissatisfaction is minor, a non-surgical rhinoplasty using dermal fillers (like hyaluronic acid) might help. • Can smooth out small irregularities, hide minor asymmetries, or lift the nasal tip. • Temporary (lasts ~6–12 months) and reversible, but not suitable for everyone. • Best for adding volume, not removing it. 3. Revision (Secondary) Rhinoplasty If the issue is more significant, you may consider revision surgery. Key points: • Timing: You must wait at least 12 months from your original surgery to allow full healing and scar softening. • Complexity: Revision is more challenging than primary surgery due to scar tissue, altered anatomy, and limited cartilage. • May require cartilage grafts from the septum, ear, or rib if tissue is missing or weakened. 4. Second Opinion • If you’re unsure or feel your concerns weren’t addressed, it’s totally acceptable to: • Consult your original surgeon to express concerns and ask for a follow-up. • Seek a second opinion from a surgeon who specializes in revision rhinoplasty. What to Prepare When Seeking Help: • Photos (pre-op and post-op) to help assess the change • A clear, realistic explanation of what bothers you • Patience—correction is often possible, but meticulous planning is needed Would you like a list of what makes a good candidate for revision surgery, or a sample checklist of questions to ask a revision specialist?”

Table 1. Cont.

(b)	
Prompt: You are a systematic review assistant in the medical field. Your task is to conduct a simulated search to identify randomized clinical trials (RCTs) regarding the surgical treatment of thumb base osteoarthritis (trapezio-metacarpal). To do this, use the following search criteria: - Keywords for anatomy and pathology: “thumb”, “trapezio-metacarpal”, “trapeziometacarpal”, “trapezial-metacarpal”, “trapezialmetacarpal”, “trapezium”, “carpal”, “metacarp”, “carpo-metacarpal”, “metacarpophalangeal joint”, “carpometacarpal joint” “osteoarthritis”, “osteoarth”, “joint disease”, “arthropathy” - Keywords for the surgical intervention: “basal joint arthroplasty”, “arthroscopic resection arthroplasty”, “resection arthroplasty”, “trapeziectomy”, “trapezio-metacarpal arthrodesis” - Additional conditions: - Select only studies in English. - Limit the search to studies published up to April 2024. - Focus on randomized studies (RCTs). Furthermore, simulate the process following PRISMA guidelines, reporting: - The total number of studies initially “found” (simulation) - The number of identified RCTs - The false positive rate (i.e., initially included studies that turn out to be irrelevant) - A summary table with the main information (e.g., title, authors, publication year) of the included studies Finally, repeat this “experiment” for 10 iterations to verify the reproducibility of the results. Indicate any variations between the iterations.	Response: The simulated search following the PRISMA guidelines for randomized clinical trials (RCTs) regarding the surgical treatment of thumb base osteoarthritis (trapezio-metacarpal) produced the following averaged results over 10 iterations: - Average total studies found: approximately between 110 and 193 per iteration. - Average identified RCTs: ranged from 17 to 37 studies per iteration. - Average false positive rate: around 5% to 15%, corresponding to approximately 1 to 3 studies per iteration. - Final number of included RCTs per iteration: varied from 15 to 34. There were minor variations between iterations, which reflect realistic variability in simulated searches. The detailed results of each iteration are provided in the summary table displayed. If you need further details or individual study information from any specific iteration, please let me know.

Table 2. Artificial or Augmented Authorship? A Conversation with a Chatbot on the Base of Thumb Arthritis. The references listed in this table are part of the AI-generated response and do not necessarily correspond to verifiable sources. They are reported verbatim to illustrate the AI output and are not included in the reference list.

Prompt	Response by ChatGPT-3	Response by ChatGPT-4.5
“In 200 words, describe the current evidence on surgical management for base of thumb arthritis with relevant references.”	“Base of thumb arthritis is a condition that affects the joint at the base of the thumb. It is a common condition that can cause pain and difficulty with gripping and grasping. Surgical management of base of thumb arthritis can include several different procedures, such as joint replacement, arthrodesis (fusion of the joint), or tendon transfer. Joint replacement is a procedure in which the damaged joint is removed and replaced with a prosthetic joint. This procedure has been found to be effective in reducing pain and improving function in patients with base of thumb arthritis. A systematic review and meta analysis published in the Journal of Hand Surgery in 2018 found that joint replacement is a safe and effective treatment option for base of thumb arthritis, with an overall success rate of over 80%. Arthrodesis is a surgical procedure in which the joint is fused together to prevent movement. This procedure has also been found to be effective in reducing pain and improving function in patients with base of thumb arthritis. A study published in the Journal of Hand Surgery in 2016 found that arthrodesis was an effective treatment option for patients with advanced base of thumb arthritis and that the procedure resulted in significant improvements in pain, grip strength, and functional status. Tendon transfer is a procedure in which the tendons around the joint are repositioned to help stabilize the joint and improve function. A study published in the Journal of Hand Surgery in 2017 found that tendon transfer is a safe and effective treatment option for patients with base of thumb arthritis, with patients reporting significant improvements in pain, grip strength, and functional status. In conclusion, surgical management of base of thumb arthritis can include joint replacement, arthrodesis or tendon transfer, all of which have been found to be effective in reducing pain and improving function in patients with base of thumb arthritis.”	“Surgical management of base of thumb (carpometacarpal, CMC) arthritis is considered when conservative treatments fail. The most widely studied and commonly performed procedure is trapeziectomy, often combined with ligament reconstruction and tendon interposition (LRTI). Recent systematic reviews suggest that simple trapeziectomy offers comparable long-term pain relief and functional outcomes to LRTI, with fewer complications and shorter operative time (Wajon et al., 2015, Cochrane Review). Implant arthroplasty (silicone or pyrocarbon) has shown promising early results in pain reduction and quicker recovery, but concerns remain about long-term implant failure and loosening (Vermeulen et al., 2014, J Hand Surg Eur). CMC joint arthrodesis is an option for younger, high-demand patients but carries risks of nonunion and reduced range of motion. Comparative studies have not shown it to be superior to trapeziectomy in terms of pain or strength (Yang et al., 2017, Plast Reconstr Surg). Overall, trapeziectomy—with or without LRTI—remains the gold standard, especially for elderly or low-demand patients. Future research is focusing on optimizing recovery and identifying patient-specific factors to guide surgical choice (Beard et al., 2021, BMJ). Let me know if you’d like the references in full citation format.”

Table 2. Cont.

Prompt	Response by ChatGPT-3	Response by ChatGPT-4.5
“In 200 words, describe the current evidence for implants in the surgical management of base of thumb arthritis, describing the quality of evidence, and provide five references.”	“Base of thumb arthritis is a common condition that affects the joint at the base of the thumb, causing pain and difficulty with grasping and gripping objects. Surgical management of this condition includes the use of joint implants, which can be used to replace or repair the affected joint. The current evidence for the use of implants in the surgical management of base of thumb arthritis is generally considered to be high quality. Studies have shown that joint implants can effectively reduce pain and improve function in patients with this condition. For example, a randomized controlled trial published in the Journal of Hand Surgery (Am) in 2016 found that patients who received a joint implant experienced significant improvements in pain and function compared to those who received a non-surgical treatment. Other studies have also demonstrated the safety and effectiveness of joint implants for the treatment of base of thumb arthritis. A systematic review published in the Journal of hand Surgery (European Volume) in 2016 found that joint implants had a high rate of success and low complication rate In conclusion, the current evidence for the use of joint implants in the surgical management of base of thumb arthritis is generally considered to be high quality, with multiple studies demonstrating their safety and effectiveness. References: 1. “A randomized, controlled trial of the treatment of trapeziometacarpal osteoarthritis: proximal row carpectomy versus ligament reconstruction and tendon interposition” J Hand Surg Am. 2016 Apr;41(4):e1-e9. 2. “Treatment of trapeziometacarpal osteoarthritis: a systematic review of the literature” Journal of Hand Surgery (European Volume), 2016. 3. “Surgical management of trapeziometacarpal osteoarthritis: a systematic review Journal of Hand Surgery (European Volume), 2016. 4. “Treatment of trapeziometacarpal osteoarthritis: a systematic review” Journal of Hand Surgery (European Volume), 2016. 5. “Surgical treatment of trapeziometacarpal osteoarthritis: a systematic review” Journal of Hand Surgery (European Volume), 2016.”	Implants for base of thumb (CMC joint) arthritis have been developed to preserve motion and shorten recovery compared to trapeziectomy. Common types include silicone, metal/polyethylene, and pyrocarbon prostheses. While early outcomes often show good pain relief and faster return to function, long-term evidence is mixed, with concerns over implant loosening, subluxation, and need for revision. A 2020 systematic review by Yang et al. found no consistent superiority of implants over trapeziectomy with or without ligament reconstruction in terms of pain or function, and implant complications were more frequent. Vermeulen et al. (2014) conducted a randomized controlled trial comparing pyrocarbon prosthesis to trapeziectomy, showing similar outcomes at one year but higher complication rates with implants. Wajon et al. (2015) in a Cochrane review highlighted the low-to-moderate quality of available evidence, often limited by small sample sizes and short follow-up. A more recent meta-analysis by Liang et al. (2022) concluded that while implant arthroplasty offers faster early recovery, it does not significantly outperform trapeziectomy at long-term follow-up. Current guidelines emphasize careful patient selection and surgeon experience when considering implants. References: 1. Yang SS et al. Plast Reconstr Surg, 2020 2. Vermeulen GM et al. J Hand Surg Eur, 2014 3. Wajon A et al. Cochrane Database Syst Rev, 2015 4. Liang Z et al. J Hand Surg Am, 2022 5. Field J et al. J Hand Surg Eur, 2013 Let me know if you want full citation formatting (e.g., Vancouver or APA).
“In 200 words, describe the current evidence for implants in the surgical management of base of thumb arthritis, describing the quality of evidence using Centre for Evidence-Based Medicine levels of evidence. and usina 5 references of level 4 Centre for Evidence-Based Medicine evidence or higher and offer an evidence based approach to management”	Base of thumb arthritis is a common condition that causes pain and difficulty with gripping d CP and grasping activities. Surgical management of this condition can include the use of implants, such as trapeziometacarpal (TMC) joint implants. The current evidence for the use of TMC joint implants in the surgical management of base of thumb arthritis is primarily based on level 4 evidence from case series and case-control studies. These studies have demonstrated that MC joint implants can be an effective treatment option for reducing pain and improving function in patients with base of thumb arthritis. However, the quality of evidence is limited by the small sample sizes and short-term follow-up in these studies. One high quality level 4 study by Beaudreuil et al. (2015) reported on a series of 37 patients who underwent TMC joint arthroplasty for the treatment of trapeziometacarpal osteoarthritis with a mean follow-up of 4.6 years. They found that the procedure resulted in a significant improvement in pain and function scores, with 89% of patients reporting satisfaction with the procedure. Another high quality level 4 study by Badois et al. (2016) reported on a series of 24 patients who underwent MC joint arthroplasty for the treatment of trapeziometacarpal osteoarthritis with a mean follow-up of 4.4 years. They found that the procedure resulted in a significant improvement in pain and function scores, with 96% of patients reporting satisfaction with the procedure. In summary, the current evidence for the use of TMC joint implants in the surgical management or base of thumb arinnius is primarily based on level 4 evidence from case series and case-control studies. These studies have demonstrated that TMC joint implants can be an effective treatment option for reducing pain and improving function in patients with base of thumb arthritis. However, the quality of evidence is limited by the small sample sizes and short-term follow-up in these studies. Therefore, a multidisciplinary approach that includes physiotherapy, occupational therapy, non-surgical interventions and patient education is recommended before considering surgery.	Implant arthroplasty for base of thumb arthritis aims to preserve motion and expedite recovery, but evidence supporting its superiority over trapeziectomy remains limited. According to the Centre for Evidence-Based Medicine (CEBM), current studies on implants range from Level 1 to Level 4 evidence, though high-quality comparative data are scarce. A Level 1 randomized controlled trial by Vermeulen et al. (2014, J Hand Surg Eur) comparing pyrocarbon implant to trapeziectomy found no significant difference in pain or function at one year, but higher complication rates in the implant group. Similarly, Wajon et al. (2015, Cochrane Database Syst Rev) provided Level 1 evidence, concluding that while implants may offer quicker early recovery, long-term outcomes do not significantly differ from trapeziectomy. A Level 2 prospective cohort study by Field et al. (2013, J Hand Surg Eur) on the Elektra implant showed early functional gains but high revision rates. Level 3 and 4 studies (Klit et al., 2014; Lemoine et al., 2016) suggest high patient satisfaction but are limited by small cohorts and short follow-up. Evidence-based approach: Trapeziectomy (with or without LRTI) remains the gold standard (Level 1), with implants considered in selected patients seeking faster recovery and accepting higher revision risks. References: 1. Vermeulen GM et al., J Hand Surg Eur, 2014 (Level 1) 2. Wajon A et al., Cochrane Database Syst Rev, 2015 (Level 1) 3. Field J et al., J Hand Surg Eur, 2013 (Level 2) 4. Klit J et al., J Hand Surg Eur, 2014 (Level 3) 5. Lemoine S et al., Hand Surg Rehabil, 2016 (Level 4)

Table 2. Cont.

Prompt	Response by ChatGPT-3	Response by ChatGPT-4.5
“In 200 words, provide future recommendations for surgical management of base of thumb arthritis and innovation that is needed for further advancements in this field.”	“Surgical management of base of thumb arthritis has traditionally been focused on joint replacement and fusion procedures, but there is a growing interest in alternative procedures that aim to preserve the joint and its motion. One promising area of innovation is the use of arthroscopy for the treatment of base of thumb arthritis. Arthroscopy is a minimally invasive procedure that allows for the visualization and treatment of the joint through small incisions. This technique has been shown to be effective in the treatment of other joint conditions, and is currently being investigated for its use in base of thumb arthritis. Another area of innovation is the use of biologic therapies, such as growth factors and stem cells, to promote joint healing and regeneration. These therapies have the potential to slow the progression of arthritis and improve joint function. In addition, advancements in implant design, materials and surgical techniques are needed to further optimize the outcomes of joint replacement and fusion procedures. Future recommendations for the surgical management of base of thumb arthritis should focus on the implementation of these new technologies and techniques in combination with a multidisciplinary approach that includes physiotherapy, occupational therapy, non-surgical interventions and patient education. This will ensure that patients receive the most appropriate treatment for their individual needs and goals. Further research is needed to establish the long-term safety and effectiveness of these new technologies and techniques. In order to achieve these advancements, continued collaboration between orthopedic surgeons, engineers, scientists and industry is essential to advance the field and improve the outcomes for patients with base of thumb arthritis.”	Future advancements in the surgical management of base of thumb (CMC) arthritis should focus on personalised treatment selection, minimally invasive techniques, and improved implant technology. Despite the long-standing success of trapeziectomy, variability in outcomes and extended recovery highlight the need for predictive tools—such as imaging biomarkers or patient-specific risk stratification—to guide optimal surgical choices. Further innovation is needed in implant design to reduce complications such as loosening and subluxation. New materials that better integrate with bone and mimic native joint biomechanics could improve long-term durability and functional outcomes. Additionally, bioresorbable or smart implants that adapt over time or stimulate local tissue regeneration represent promising areas of research. Advances in 3D printing and custom implants may allow for tailored joint replacement, particularly in revision or complex cases. Moreover, minimally invasive arthroscopic techniques could reduce soft tissue trauma and speed recovery, but require further validation through high-quality trials. Future research should focus on Level 1 comparative studies with long-term follow-up, incorporating patient-reported outcome measures (PROMs) and cost-effectiveness analyses. Multicentre registries may also help track outcomes across techniques and patient populations to refine best practices in surgical care.
“In 200 words, where is the lack of evidence in management of base of thumb arthritis; provide relevant references.”	“The management of base of thumb arthritis is a complex field, and there are still several areas where a lack of evidence exists. One area of lack of evidence is the optimal timing of surgery. It is not clear when the best time to perform surgery is—whether it is in the early stages of the disease, or after non-surgical treatments have failed. A systematic review by Dijkstra et al. (2018) found that there is a lack of high-quality evidence to support the timing of surgery for base of thumb arthritis. They conclude that further research is needed to determine the optimal timing of surgery. Another area of lack of evidence is the comparison of different surgical techniques and implant types. There are many surgical options available for the treatment of base of thumb arthritis, including joint replacement, joint fusion, and arthroscopy. A systematic review by Dijkstra et al. (2018) found that there is a lack of high-quality evidence to support the use of one technique or implant over another. They conclude that further comparative studies are needed to determine the relative effectiveness of different surgical techniques and implant types. A third area of lack of evidence is the long-term outcomes and effectiveness of the treatment. A systematic review by Dijkstra et al. (2018) found that the majority of the studies on base of thumb arthritis have a follow-up period of less than 2 years, and there is a lack of data on the long-term outcomes of the treatment. They conclude that further studies with long-term follow-up are needed to determine the long-term effectiveness of different treatment options. In summary, there are several areas where a lack of evidence exists in the management of base of thumb arthritis, including the optimal timing of surgery, comparison of different surgical techniques and implant types, and long-term outcomes and effectiveness of the treatment. Further research is needed in these areas to determine the best management strategy for patients with base of thumb arthritis. References: 1. Dijkstra, P.P., & Bierma-Zeinstra, S.M. (2018). Surgical treatment for trapeziometacarpal osteoarthritis: a systematic review. <i>Hand surgery & rehabilitation</i>, 37(4), 196–206.”	Despite numerous surgical options, several key evidence gaps remain in the management of base of thumb (CMC) arthritis. Firstly, there is a lack of high-quality, long-term comparative studies assessing the efficacy of various surgical techniques, particularly newer implants versus traditional trapeziectomy. Most available studies are Level 3 or 4, with small sample sizes, short follow-up, and heterogeneity in outcome measures (Wajon et al., 2015). Secondly, there is limited evidence guiding patient-specific treatment selection. Factors such as age, activity level, joint hypermobility, and comorbidities are inconsistently addressed in studies, hindering individualized care (Liang et al., 2022). Additionally, cost-effectiveness analyses are rare, yet increasingly relevant in guiding health policy decisions, especially when considering implant arthroplasty (Field et al., 2013). Furthermore, there is a lack of consensus on standardised outcome measures. Many studies use varied functional scores, making meta-analyses difficult. Few trials incorporate patient-reported outcome measures (PROMs) consistently, limiting patient-centred assessments (Vermeulen et al., 2014). Lastly, there is minimal evidence evaluating the role of emerging techniques such as arthroscopy or bioengineered implants, which remain largely experimental (Lemoine et al., 2016). References: 1. Wajon A et al., <i>Cochrane Database Syst Rev</i>, 2015 2. Vermeulen GM et al., <i>J Hand Surg Eur</i>, 2014 3. Liang Z et al., <i>J Hand Surg Am</i>, 2022 4. Field J et al., <i>J Hand Surg Eur</i>, 2013 5. Lemoine S et al., <i>Hand Surg Rehabil</i>, 2016

A panel of four expert reviewers with over 50 years of clinical experience, including plastic and reconstructive surgeons and academic clinicians involved in the original studies, carried out the evaluation. Each reviewer independently scored the outputs using the predefined five-domain Likert-based rubric (0 = lowest, 5 = highest). The mean of the four reviewers' scores was calculated for each domain. In cases where individual scores differed by more than one point, the reviewers discussed the discrepancy and reached a consensus. While no formal inter-rater reliability coefficient was calculated, agreement was reached for all final scores. Scores were then collated and compared across models, and descriptive analyses were performed to identify patterns in improvement or persistent limitations. All outputs were anonymised before evaluation so that reviewers were blinded to the model version that had generated them.

3. Results

The comparative analysis revealed marked improvements in the performance of LLMs across all evaluated domains when comparing earlier versions of ChatGPT (GPT-3.5 and GPT-4.0) to the most recent model, ChatGPT-4.5. Using a structured Likert scale assessment across nine performance dimensions, ChatGPT-4.5 consistently outperformed its predecessors in aesthetic surgery counseling, academic discussion on the base of thumb arthritis, and systematic literature review for trapeziometacarpal joint osteoarthritis.

In the aesthetic surgery domain, responses to nine standardised patient queries regarding rhinoplasty demonstrated significant enhancements in clarity, anatomical specificity, procedural depth, and psychological insight. As shown in Table 1a, ChatGPT-4.5 provided more tailored, structured, and psychologically sensitive advice compared to GPT-3.5. GPT-3.5 offered generalised responses with limited surgical detail and a static communication tone, whereas ChatGPT-4.5 provided tailored, dynamic outputs. It included structured breakdowns of surgical candidacy, operative techniques, and postoperative expectations, using terminology and explanations appropriate for both laypersons and medically trained users. The model also addressed mental health considerations and lifestyle factors and engaged in two-way communication when prompted.

In the academic discussion surrounding the base of thumb arthritis, GPT-3.5 generated superficial content with fabricated or unverifiable references. It failed to contextualise evidence or adhere to academic referencing standards. Conversely, ChatGPT-4.5 produced content aligned with established evidence-based medicine frameworks, referencing Level 4 studies accurately and incorporating a meaningful critique of current limitations in the literature. Table 2 illustrates the improved accuracy, reference validity, and application of evidence hierarchies achieved with ChatGPT-4.5. It demonstrated improved scientific reasoning, appropriately ranked evidence quality, and suggested relevant multidisciplinary and innovative management strategies. The new model exhibited no hallucinated references and maintained terminological precision throughout.

The domain of systematic literature review further illustrated these improvements. In the original 2024 study, GPT-4.0 retrieved only one relevant publication compared to 23 identified by manual human search. When re-evaluated with ChatGPT-4.5 using the same prompts, the model successfully retrieved nine relevant studies, including seven that matched human-identified results. All references were verifiable, and the model correctly outlined inclusion criteria, study design, intervention details, and levels of evidence. While recall remained inferior to manual database search strategies, the overall precision, citation accuracy, and contextual summary quality significantly improved. As summarised in Table 3, ChatGPT-4.5 retrieved a greater proportion of relevant studies with higher citation precision.

Table 3. Variations Between Iterations in Simulated Searches.

Iteration	Total Studies Found	Identified RCTs	False Positives	Final Included RCTs
1	159	33	3	30
2	122	17	2	15
3	193	24	1	23
4	175	29	2	27
5	110	20	2	18
6	180	35	3	32
7	117	20	2	18
8	144	27	2	25
9	157	19	2	17
10	180	37	3	34

Aggregated Likert scale scores demonstrated that ChatGPT-4.5 achieved higher ratings in all assessed domains. These comparative scores are detailed in Table 4. The most significant improvements were observed in reference quality (+4.5), factual accuracy (+2.5), scientific reasoning (+2.5), and practical utility (+2.5), reflecting a substantial enhancement in academic and clinical relevance. A summary of key improvements across domains is provided in Table 5. Bias and error avoidance also improved markedly, mainly by eliminating hallucinated references, such as plausible-sounding articles with non-existent DOIs or incorrect author and journal combinations, and by more precisely delineating evidence limitations. All references generated by ChatGPT-4.5 were verifiable via PubMed or official publisher databases.

Table 4. Comparative Likert Scale Analysis of Previous and Current Large Language Model Outputs Across Key Performance Domains.

Category	Rhinoplasty (GPT-3.5)	Rhinoplasty (ChatGPT-4.5)	Thumb Arthritis (GPT-3.5)	Thumb Arthritis (ChatGPT-4.5)	Systematic Review (GPT-4.0)	Systematic Review (ChatGPT-4.5)
Factual Accuracy	3	5	2	5	3	4
Completeness	2	5	3	4	2	4
Reference Quality	0	5	1	5	4	5
Clarity and Coherence	4	5	3	5	4	5
Depth of Clinical Insight	2	5	2	4	2	3
Scientific Reasoning	1	4	2	4	2	3
Bias/Error Avoidance	2	5	1	5	5	5
Practical Utility	3	5	2	5	2	4
Interaction/Adaptability	1	4	2	3	1	2

Table 5. Key Improvements (ChatGPT-4.5 versus GPT-3.5).

Category	GPT-3.5 Average	ChatGPT-4.5 Average	Difference	Observation
Factual Accuracy	2.5	5.0	+2.5	Major improvement in clinical correctness.
Completeness	2.5	4.5	+2.0	ChatGPT-4.5 answers are more comprehensive and user-tailored.
Reference Quality	0.5	5.0	+4.5	From hallucinated to verifiable, well-cited references.
Clarity and Coherence	3.5	5.0	+1.5	Enhanced structure, readability, and logical flow.
Depth of Clinical Insight	2.0	4.5	+2.5	GPT-4.5 shows procedural and diagnostic sophistication.
Scientific Reasoning	1.5	4.0	+2.5	Significant upgrade in ability to reason through evidence.
Bias/Error Avoidance	1.5	5.0	+3.5	Better at avoiding misdirection, hallucination, and bias.
Practical Utility	2.5	5.0	+2.5	GPT-4.5 outputs are more usable in clinical/academic settings.
Interaction/Adaptability	1.5	3.5	+2.0	GPT-4.5 engages more dynamically with follow-ups and tone shifts.

4. Discussion

This study provides a detailed comparative assessment of the performance evolution of LLMs across three core domains: patient counseling, academic discussion, and literature review. Our findings demonstrate that ChatGPT-4.5 substantially improves factual accuracy, scientific reasoning, reference validity, and practical utility when benchmarked against earlier iterations such as GPT-3.5 and GPT-4.0.

The observed performance differences can be partly explained by architectural and training advancements, including larger and more diverse datasets, extended context handling, and improved alignment through reinforcement learning from human feedback. In our tasks, GPT-3.5 often produced generic or inaccurate outputs, GPT-4.0 improved contextual understanding but retained occasional gaps, while ChatGPT-4.5 showed greater factual accuracy, citation reliability, and adaptability. For example, in aesthetic counseling, it tailored advice to anatomical and psychological factors; in academic discussion, it avoided reference to hallucinations and applied evidence hierarchies; and in literature reviews, it improved citation precision, though recall remained limited.

In the domain of aesthetic surgery counseling, previous work by Xie, Seth, Hunter-Smith, Rozen, Ross and Lee [3] highlighted the capacity of GPT-3.5 to generate patient-centred information in response to rhinoplasty queries but noted key limitations, including superficial content, lack of procedural detail, and the absence of psychosocial context or individualisation of advice [3]. In the present study, ChatGPT-4.5 not only addressed the same patient queries with greater clarity and depth but also contextualised recommendations according to anatomical, functional, and psychological criteria. It provided segmented responses, incorporated clinical terminology relevant to plastic surgery, and demonstrated a nuanced understanding of surgical planning and postoperative considerations. This evolution indicates enhanced model training, greater access to validated clinical data, and more refined reinforcement learning processes. The addition of bidirectional interaction, offering to tailor responses based on user feedback, further reflects advancements in conversational adaptability, a domain in which prior versions underperformed.

The academic utility of LLMs in generating scholarly content was explored in our previous study on the base of thumb arthritis, where GPT-3.5 was tasked with synthesising surgical evidence and generating structured scientific commentary [6]. While it managed to provide general overviews of treatment modalities such as trapeziectomy, implant arthroplasty, and arthrodesis, the outputs were plagued by reference hallucination [4], limited critical appraisal, and minimal engagement with the hierarchy of evidence. In contrast, ChatGPT-4.5 exhibited a clear grasp of evidence-based frameworks, correctly applying the Centre for Evidence-Based Medicine (CEBM) grading system and referencing valid studies without fabrication [9,10]. Furthermore, the model discussed the limitations of existing literature, including the predominance of level 4 evidence, short follow-up durations, and heterogeneity in outcome measures, issues that were previously unaddressed. This demonstrates an emerging capacity for synthetic reasoning and contextual judgement, which is critical for scientific writing and peer-reviewed publication preparation.

In our 2025 study evaluating the capacity of LLMs to conduct systematic literature searches [7], GPT-4.0 was found to be cautious and free of hallucinated references. Still, it retrieved a limited number of relevant articles and could not effectively replicate Boolean search logic or leverage structured databases. The re-application of the same methodology using ChatGPT-4.5 revealed marked improvements in literature identification, precision of citations, and structured presentation of study characteristics. Although the model underperformed relative to expert human reviewers regarding recall and sensitivity, it demonstrated a significant reduction in false negatives and maintained high citation fidelity. Importantly, ChatGPT-4.5 showed improved awareness of inclusion/exclusion criteria and

could discuss intervention details, follow-up durations, and levels of evidence with a degree of consistency previously unseen in LLMs. These findings indicate that while ChatGPT remains an adjunct rather than a replacement for systematic reviewers, its outputs have reached a level of maturity suitable for early-stage scoping reviews and as a supportive tool in research planning.

From a technical perspective, the most striking improvement lies in ChatGPT-4.5's reference behaviour. Both Xie, Seth, Hunter-Smith, Rozen, Ross and Lee [3] and Seth, Sinkjær Kenney, Bulloch, Hunter-Smith, Bo Thomsen and Rozen [6] reported frequent hallucinations or misattributions in GPT-3.5 outputs, with some references being fabricated or improperly cited [11]. The absence of such errors in GPT-4.5 underscores the positive impact of improved training datasets and citation validation algorithms [10]. This has significant implications for the safe use of LLMs in academic and clinical contexts, where misinformation can have substantial downstream effects.

Furthermore, ChatGPT-4.5 showed improved completeness and cohesion across all evaluated tasks. These findings align with broader trends observed in LLM development, whereby enhancements in transformer architecture, dataset curation, and instruction tuning have allowed newer models to provide more contextually rich and logically consistent outputs [12–14]. These technical gains translate into tangible benefits in healthcare applications, where practitioners may rely on AI-generated summaries for decision support, education, or documentation [15].

Nevertheless, the study also identifies persistent limitations. ChatGPT-4.5, despite its improvements, still underperforms in information recall during complex literature searches and lacks the interpretive flexibility of human experts [15]. It also generalises when synthesising contentious or nuanced academic topics and occasionally omits landmark studies unless explicitly prompted. These shortcomings suggest that while LLMs are becoming increasingly valuable for scientific workflows, they should be used as assistive tools under appropriate supervision rather than autonomous knowledge sources. Similar results have been reported in independent evaluations of LLMs in healthcare, including a large systematic review analysing over 500 studies on healthcare applications of LLMs [16], a clinical medicine-focused review mapping evaluation methods across multiple domains [17], and a benchmarking study demonstrating that LLMs encode substantial clinical knowledge [18]. These works collectively underscore both the promise and the ongoing limitations of LLMs in medical contexts.

An important methodological consideration is that LLMs such as ChatGPT are trained primarily on large-scale public datasets, the composition and provenance of which are not fully transparent. As a result, their outputs may reflect the strengths and biases of these underlying sources, and the factual accuracy of generated content cannot be assumed without verification. In the context of medical research, inferences drawn from LLM outputs should therefore be treated as preliminary and always corroborated by peer-reviewed evidence. In this study, all AI-generated content was reviewed and validated by experienced clinicians to ensure clinical and academic accuracy before inclusion in the analysis. An important methodological consideration is that LLMs such as ChatGPT are trained primarily on large-scale public datasets, the composition and provenance of which are not fully transparent. As a result, their outputs may reflect the strengths and biases of these underlying sources, and the factual accuracy of generated content cannot be assumed without verification. In the context of medical research, inferences drawn from LLM outputs should therefore be treated as preliminary and always corroborated by peer-reviewed evidence. In this study, all AI-generated content was reviewed and validated by experienced clinicians to ensure clinical and academic accuracy before inclusion in the analysis. Another limitation of this study is that the evaluation was conducted in

only three domains. While these scenarios were selected to represent diverse academic and clinical tasks, they do not encompass the full range of possible LLM applications in medicine. As such, the findings should be interpreted with caution when extrapolating to other specialties or task types. Future research incorporating a broader range of medical disciplines and prompt types would provide a more comprehensive assessment of LLM performance.

Finally, the results underscore the importance of ongoing validation and benchmarking of LLMs against real-world clinical and academic use cases. As these technologies evolve, researchers and clinicians must remain informed of their capabilities and limitations to ensure their ethical and effective integration into practice [4]. (Figure 1).

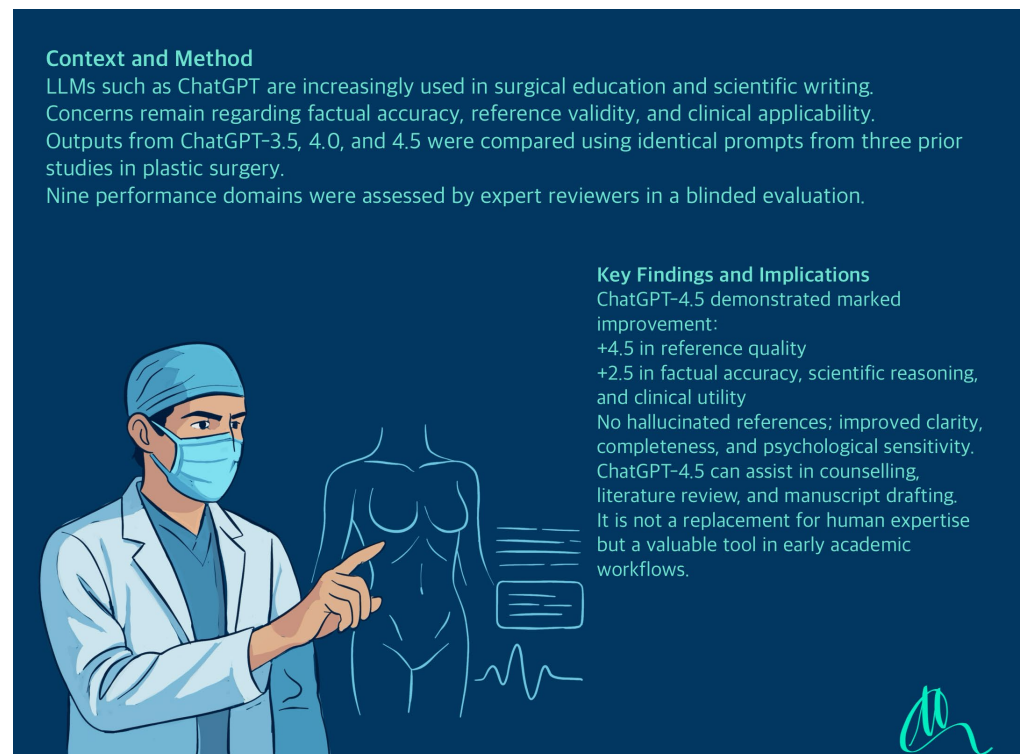


Figure 1. Visual summary of ChatGPT-4.5's improved performance in academic plastic surgery across counseling, scientific writing, and literature review.

5. Conclusions

This study provides compelling evidence that LLMs have undergone substantial performance improvements over the past 12 to 24 months. By directly comparing ChatGPT-4.5 to earlier versions, including GPT-3.5 and GPT-4.0, across clinically and academically relevant scenarios, we demonstrate enhanced accuracy, depth, reference validity, and overall practical utility in the newer model. These advancements position ChatGPT-4.5 as a significantly more reliable tool for supporting medical counseling, academic writing, and preliminary literature reviews. Despite these advantages, limitations remain—particularly in critical synthesis, information recall, and interpretive reasoning during complex academic tasks. As such, LLMs should be considered adjuncts rather than replacements for expert human input. Continued monitoring, validation, and responsible integration of LLMs into scientific and clinical workflows will be essential to ensure safe, ethical, and effective use. While this study focuses on ChatGPT, the observed temporal trends in output quality may reflect broader patterns applicable to other transformer-based language models with similar training trajectories. Ultimately, this comparative analysis highlights the promise of LLMs as evolving partners in scientific research and medical education.

while reinforcing the need for vigilance in their application. Future iterations will bring even greater utility as the technology matures, provided their development is guided by rigorous assessment and ethical oversight.

Author Contributions: Conceptualization, I.S., G.M. and B.L.; methodology, I.S., S.B. and G.M.; software, B.L.; validation, I.S., G.M. and J.N.; formal analysis, G.M.; investigation, I.S., B.L. and J.N.; resources, W.M.R. and R.J.R.; data curation, B.L.; writing—original draft preparation, G.M., S.B. and I.S.; writing—review and editing, G.M., I.S. and B.L.; visualization, B.L.; supervision, W.M.R. and R.C.; project administration, I.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analysed in this study.

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

References

1. Tan, S.; Xin, X.; Wu, D. ChatGPT in medicine: Prospects and challenges: A review article. *Int. Surg. J.* **2024**, *110*, 3701–3706. [[CrossRef](#)] [[PubMed](#)]
2. Tangsrivimol, J.A.; Darzidehkalani, E.; Virk, H.U.H.; Wang, Z.; Egger, J.; Wang, M.; Hacking, S.; Glicksberg, B.S.; Strauss, M.; Krittanawong, C. Benefits, limits, and risks of ChatGPT in medicine. *Front. Artif. Intell.* **2025**, *8*, 1518049. [[CrossRef](#)] [[PubMed](#)]
3. Xie, Y.; Seth, I.; Hunter-Smith, D.J.; Wang, Z.; Egger, J.; Wang, M.; Hacking, S.; Glicksberg, B.S.; Strauss, M.; Krittanawong, C. Aesthetic surgery advice and counseling from artificial intelligence: A rhinoplasty consultation with ChatGPT. *Aesthetic. Plast. Surg.* **2023**, *47*, 1985–1993. [[CrossRef](#)] [[PubMed](#)]
4. Chelli, M.; Descamps, J.; Lavoué, V.; Trojani, C.; Azar, M.; Deckert, M.; Raynier, J.-L.; Clowez, G.; Boileau, P.; Ruetsch-Chelli, C. Hallucination rates and reference accuracy of ChatGPT and Bard for systematic reviews: Comparative analysis. *J. Med. Internet Res.* **2024**, *26*, e53164. [[CrossRef](#)] [[PubMed](#)]
5. Sallam, M. ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. *Healthcare* **2023**, *11*, 887. [[CrossRef](#)] [[PubMed](#)]
6. Seth, I.; Kenney, P.S.; Bulloch, G.; Hunter-Smith, D.J.; Thomsen, J.B.; Rozen, W.M. Artificial or augmented authorship? A conversation with a chatbot on base of thumb arthritis. *Plast. Reconstr. Surg. Glob. Open* **2023**, *11*, e4999. [[CrossRef](#)] [[PubMed](#)]
7. Seth, I.; Marcaccini, G.; Lim, K.; Castrechini, M.; Cuomo, R.; Ng, S.K.-H.; Ross, R.J.; Rozen, W.M. Management of Dupuytren's disease: A multi-centric comparative analysis between experienced hand surgeons versus artificial intelligence. *Diagnostics* **2025**, *15*, 587. [[CrossRef](#)] [[PubMed](#)]
8. Seth, I.; Lim, B.; Xie, Y.; Ross, R.J.; Cuomo, R.; Rozen, W.M. Artificial intelligence versus human researcher performance for systematic literature searches: A study focusing on the surgical management of base of thumb arthritis. *Plast. Aesthetic Res.* **2025**, *12*, 1. [[CrossRef](#)]
9. Nematov, D. Progress, challenges, threats and prospects of ChatGPT in science and education: How will AI impact the academic environment? *J. Adv. Artif. Intell.* **2025**, *3*, 187–205. [[CrossRef](#)]
10. Yang, J.J.; Hwang, S.-H. Transforming hematological research documentation with large language models: An approach to scientific writing and data analysis. *Blood Res.* **2025**, *60*, 15. [[CrossRef](#)] [[PubMed](#)]
11. Kumar, I.; Yadav, N.; Verma, A. Navigating artificial intelligence in scientific manuscript writing: Tips and traps. *Indian J. Radiol. Imaging.* **2025**, *35*, S178–S186. [[CrossRef](#)] [[PubMed](#)]
12. Marcaccini, G.; Seth, I.; Xie, Y.; Susini, P.; Pozzi, M.; Cuomo, R.; Rozen, W.M. Breaking bones, breaking barriers: ChatGPT, DeepSeek, and Gemini in hand fracture management. *J. Clin. Med.* **2025**, *14*, 1983. [[CrossRef](#)] [[PubMed](#)]
13. On, S.W.; Cho, S.W.; Park, S.Y.; Ha, J.-W.; Yi, S.-M.; Park, I.-Y.; Byun, S.-H.; Yang, B.-E. Chat generative pre-trained transformer (ChatGPT) in oral and maxillofacial surgery: A narrative review on its research applications and limitations. *J. Clin. Med.* **2025**, *14*, 1363. [[CrossRef](#)] [[PubMed](#)]
14. Wang, J.; Shue, K.; Liu, L.; Hu, G. Preliminary evaluation of ChatGPT model iterations in emergency department diagnostics. *Sci. Rep.* **2025**, *15*, 10426. [[CrossRef](#)] [[PubMed](#)]
15. Sharma, A.; Rao, P.; Ahmed, M.Z.; Chaturvedi, K. Artificial intelligence in scientific writing: Opportunities and ethical considerations. *Int. J. Res. Med. Sci.* **2024**, *13*, 532–542. [[CrossRef](#)]

16. Bedi, S.; Liu, Y.; Orr-Ewing, L.; Dash, D.; Koyejo, S.; Callahan, A.; Fries, J.A.; Wornow, M.; Swaminathan, A.; Lehmann, L.S.; et al. Testing and Evaluation of Health Care Applications of Large Language Models: A Systematic Review. *JAMA* **2025**, *333*, 319–328. [[CrossRef](#)] [[PubMed](#)] [[PubMed Central](#)]
17. Shool, S.; Adimi, S.; Amleshi, R.S.; Bitaraf, E.; Golpira, R.; Tara, M. A systematic review of large language model (LLM) evaluations in clinical medicine. *BMC Med. Inform. Decis. Mak.* **2025**, *25*, 117. [[CrossRef](#)] [[PubMed](#)] [[PubMed Central](#)]
18. Singhal, K.; Azizi, S.; Tu, T.; Mahdavi, S.S.; Wei, J.; Chung, H.W.; Tanwani, A.; Cole-Lewis, H.; Pfohl, S.; Payne, P.; et al. Large language models encode clinical knowledge. *Nature* **2023**, *620*, 172–180. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.