Journal Pre-proof

Generative Artificial Intelligence as a Research Partner in Orthopaedics: State of the Art

Dr. Ricardo Bastos, Prof., Américo Dias, Dr. Paulo Amado, Prof., Dr. Philippe Neyret, Prof., Dr. João Espregueira-Mendes, Prof., Sthefan Gabriel Berwanger

PII: S2059-7754(26)00062-3

DOI: https://doi.org/10.1016/j.jisako.2026.101126

Reference: JISAKO 101126

To appear in: *Journal of ISAKOS*

Received Date: 10 January 2026

Revised Date: 28 March 2026

Accepted Date: 23 April 2026

Please cite this article as: Bastos R, Dias A, Amado P, Neyret P, Espregueira-Mendes J, Berwanger SG, Generative Artificial Intelligence as a Research Partner in Orthopaedics: State of the Art, *Journal of ISAKOS*, https://doi.org/10.1016/j.jisako.2026.101126.

Title: Generative Artificial Intelligence as a Research Partner in Orthopaedics: State of the Art

Author names and affiliations:

Prof. Dr. Ricardo Bastos:
Trofa Saúde Penafiel Hospital, Penafiel, Portugal.
Fluminense Federal University, Niterói, Rio de Janeiro, Brazil
PBS – Porto Business School, Porto, Portugal

Américo Dias:
PBS – Porto Business School, Porto, Portugal

Prof. Dr. Paulo Amado:
Coordinator of the Foot and Ankle Unit at Hospital Lusíadas Porto.Director of the Surgical Center at Hospital São Martinho, Valongo, Portugal.
Medical Director of Clínica Médica da Foz.
Past President of the Portuguese Foot and Ankle Society.
Board Member of the Portuguese Society of Orthopaedics and Traumatology (SPOT).

Prof. Dr. Philippe Neyret:
Infirmerie Protestante, Caluire, Lyon, France

Prof. Dr. João Espregueira-Mendes:
Clínica Espregueira-FIFA Medical Centre of Excellence, Porto, Portugal.
Dom Henrique Research Centre, Porto, Portugal.
School of Medicine, Minho University, Braga, Portugal.
ICVS/3B's–PT Government Associate Laboratory, Braga/Guimarães, Portugal.
3B's Research Group-Biomaterials, Biodegradables and Biomimetics, University of Minho, Headquarters of the European Institute of Excellence on Tissue Engineering and Regenerative Medicine, Barco, Guimarães, Portugal.

Sthefan Gabriel Berwanger:
Inesc TEC – Institute for Systems and Computer Engineering, Technology and Science, Porto, Portugal

Corresponding author:

Ricardo Bastos
rpsbastos1@gmail.com

ORCID: 0000-0002-8207-6708

Abstract

The integration of Artificial Intelligence (AI) into medicine has progressed from discriminative models to Generative AI (GenAI), which can synthesize novel content. For orthopaedic surgeons, scientific publication remains a vital marker of academic success but is often constrained by clinical workload. This review proposes a structured, practical framework to help orthopaedists effectively harness AI tools, transitioning from opaque, "black box" generation to grounded, verifiable research assistance through Retrieval-Augmented Generation (RAG).

A PubMed search was conducted to explore the application of GenAI in the context of orthopaedic scientific research. An interactive review with experts in GenAI was also conducted, from which the proposed structure was developed. From this synthesis, a three-phase workflow is proposed: (1) Evidence selection using semantic discovery systems to identify and map relevant literature beyond keyword matching; (2) Data extraction and synthesis employing RAG-based systems to anchor AI responses to verified PDF sources, thereby minimizing hallucinations; and (3) Drafting and refining using Large Language Models (LLMs) for structured composition, linguistic clarity, and iterative manuscript improvement.

The workflow integrates platform features to enhance efficiency, accuracy, and accessibility in orthopaedic research. When applied within a controlled, evidence-grounded environment, these systems can automate literature synthesis, expedite data extraction, and assist with scientific writing, while preserving authorial intent and accountability.

However, challenges remain. Risks include algorithmic bias, "hallucinations", privacy concerns, and ethical issues related to authorship. Despite these limitations, AI represents a paradigm shift in orthopaedic scholarship, functioning as a cognitive exoskeleton that augments rather than replaces human expertise. With vigilant human oversight and adherence to journal ethics, orthopaedic surgeons can leverage AI to enhance research productivity, reproducibility, and quality while upholding the highest standards of scientific integrity.

Keywords: Artificial intelligence; Generative AI; Scientific writing; Orthopaedic research; Large language models; Retrieval-Augmented Generation

What are the new findings?

- A three-phase workflow integrating semantic discovery, Retrieval-Augmented Generation, and Large Language Models enables structured, evidence-grounded research assistance in orthopaedic scientific writing
- Grounding artificial intelligence outputs in clinician-supplied documents through closed-corpus retrieval reduces fabrication and improves traceability compared with unconstrained model generation
- Structured prompt engineering with explicit task definitions, input constraints, and output formats improves reproducibility and interpretive control in data extraction and manuscript drafting
- Systematic human verification, audit trail documentation, and transparent disclosure of artificial intelligence use are essential safeguards to preserve scientific integrity and authorship accountability

Current concepts:

• Generative AI tools can function as research assistants across literature discovery, data extraction, and manuscript drafting when grounded in verifiable sources
• Retrieval-Augmented Generation, an approach that couples LLMs with retrieval from user-supplied sources by anchoring AI outputs to user-supplied, peer-reviewed documents rather than relying on unconstrained model generation, a technique that mitigates but does not eliminate hallucinations
• Effective prompt engineering and section-by-section drafting preserve authorial control and scientific accountability throughout the writing process
• Risks, including hallucinations, bias, data privacy violations, and narrative homogenization, require systematic human verification and transparent disclosure at every stage

Future perspectives:
• Agentic AI systems capable of autonomous multi-step task orchestration may further extend workflow automation in orthopaedic research under human oversight
• Standardized evaluation criteria and reporting conventions for AI-assisted research are needed to ensure comparability across studies and consistent editorial assessment
• Development of curated registries of GDPR- and HIPAA-compliant AI tools for research workflows would support safer and more equitable adoption
• Empirical validation of AI-assisted workflows and governance frameworks for responsible integration of agentic AI into academic practice remains a critical research priority

Text Box 1: Key articles

1. Oettl FC, Pruneski JA, Zsidai B, et al. Is orthopaedics entering the age of generative AI?—A narrative review of current applications, challenges, and future directions. *Knee Surg Sports Traumatol Arthrosc*. 2025.
2. Slawaska-Eng D, Bourgeault-Gagnon Y, Cohen D, et al. ChatGPT-3.5 and -4 provide mostly accurate information when answering patients' questions relating to femoroacetabular impingement syndrome and arthroscopic hip surgery. *JISAKOS*. 2025;10:100376.
3. Hand C, Bohn C, Tannir S, et al. American Academy of Orthopaedic Surgeons OrthoInfo provides more readable information regarding rotator cuff injury than ChatGPT. *JISAKOS*. 2025;12:100841.
4. Lack BT, Mouhawasse E, Childers JT, et al. Can ChatGPT answer patient questions regarding reverse shoulder arthroplasty? *JISAKOS*. 2024;9:100323.
5. Lewis P, Perez E, Piktus A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. *Adv Neural Inf Process Syst*. 2020;33:9459–9474.
6. Kacena MA, Plotkin LI, Fehrenbacher JC. The Use of Artificial Intelligence in Writing Scientific Review Articles. *Curr Osteoporos Rep*. 2024;22:115–121.
7. Alkaissi H, McFarlane SI. Artificial hallucinations in ChatGPT: implications in scientific writing. *Cureus*. 2023;15(2):e35179.
8. Silva JCMC, Gouveia RP, Zielinski KMC, et al. AI-Assisted Tools for Scientific Review Writing: Opportunities and Cautions. *ACS Appl Mater Interfaces*. 2025;17:47795-47805.
9. Dahmen J, Kayaalp ME, Ollivier M, et al. Artificial intelligence bot ChatGPT in medical research: the potential game changer as a double-edged sword. *Knee Surg Sports Traumatol Arthrosc*. 2023;31:1187–1189.

10. Editors IC of MJ. Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. Updated April 2025. Published online 2025. https://www.icmje.org/icmje-recommendations.pdf

Text Box 2: Validated outcome measures and classifications

• *Not applicable: methodological framework article, rather than a disease-specific outcome review.*

Text Box 3: Key issues of patient selection

• *Not applicable: Article addresses researchers, rather than clinical patient selection.*

Text Box 4: Essential and/or typical features of Artificial Intelligence Tools

• **Generative Artificial Intelligence:** Is a class of artificial intelligence systems that generate novel content, such as text, images, audio, or code, through probabilistic modeling of patterns learned from large datasets, rather than by reproducing specific inputs.

• **Large Language Models (LLMs):** Probabilistic models trained to generate text sequences. Effective for outlining, summarization, and linguistic refinement, but lacking inherent awareness of source validity or evidentiary hierarchy and therefore prone to hallucinations without external grounding.

• **Retrieval-Augmented Generation (RAG):** An approach that couples LLMs with retrieval from user-supplied sources, constraining outputs to verifiable content. RAG improves traceability and reduces fabrication, but still does not replace human verification.

• **Semantic Discovery Systems:** AI-assisted retrieval tools that use semantic similarity and citation relationships to identify relevant literature beyond keyword matching, supporting discovery and citation mapping rather than evidence synthesis.

• **Prompt Engineering:** The deliberate design of structured inputs and instructions to guide AI behavior toward accurate, relevant, and reproducible results. In orthopaedic research, effective prompt engineering defines the task, context, and output format

• **Agentic AI:** Systems designed to autonomously perform multi-step tasks by dynamically chaining prompts, retrieving data, and executing actions across connected tools or datasets. These systems extend traditional GenAI by adding reasoning and task orchestration, thereby enabling partial workflow automation under human oversight.

Text Box 5: Tips and tricks

• **Citation-network snowballing:** Use semantic discovery systems to explore forward and backward citation networks, enabling identification of seminal studies, influential methodologies, and emerging evidence beyond keyword-based searches.

• **Closed-corpus grounding:** When extracting or synthesizing factual information, restrict AI systems to a curated set of user-supplied documents to anchor outputs to verifiable sources and minimize the likelihood of hallucinations.

• **Structured prompting for constrained tasks:** Define tasks, inputs, constraints, and output formats explicitly when using Generative Artificial Intelligence for extraction, synthesis, or critique to improve reproducibility, transparency, and interpretive control.

• **Section-level, author-driven drafting:** Draft manuscripts incrementally by section, following conventional scholarly structures, such as IMRAD (Introduction, Methods, Results, Discussion), with substantive content authored by the surgeon and GenAI used for critique and refinement.

• **Targeted linguistic refinement:** Apply Generative Artificial Intelligence selectively to improve clarity, flow, and academic language, particularly for non-native English speakers, while preserving scientific meaning and retaining full author accountability.

Text Box 6: Major Pitfalls of Artificial Intelligence in Research

• **Hallucinations:** The generation of fabricated or unsupported facts, data, or citations by Large Language Models, particularly when outputs are not grounded in verifiable sources.

• **Algorithmic and representational bias:** AI-generated content may reflect or amplify biases embedded in training data, potentially influencing the framing of disease prevalence, outcomes, or clinical relevance.

• **Narrative homogenization:** Overreliance on standardized generative models may erode authorial voice, critical reasoning, and interpretive nuance, leading to uniformity in scientific discourse.

• **Data privacy and confidentiality risks:** Entering patient-level data, unpublished manuscripts, or confidential materials into non-governed AI platforms may violate privacy regulations, such as HIPAA, GDPR, and intellectual property protections.

• **Authorship and accountability limitations:** AI systems cannot meet authorship criteria, as they lack responsibility for study design, interpretation, and final approval; full accountability remains with human authors.

INTRODUCTION

Orthopaedic surgery is a specialty intrinsically linked to technological innovation, ranging from robotic navigation to 3D-printed implants.[1,2] However, the field is currently witnessing a fundamental transformation with the entry of the “generative era” of Artificial Intelligence (AI). Generative Artificial Intelligence (GenAI) is a subset of AI that focuses on producing novel content, such as text, images, audio, or code, by learning patterns from existing data rather than directly replicating it.[3,4]

Another related concept is Large Language Models (LLMs), which are a subset of GenAI models and are probabilistic language models trained to predict and generate coherent text sequences. They are highly effective for outlining, summarization, and linguistic refinement but lack an inherent understanding of source validity or evidentiary hierarchy, making them prone to factual errors and hallucinations without external grounding.[5] LLMs are notably powerful because they are pre-trained on enormous quantities of text, often internet-scale data, to comprehend and generate human-like text.[3]

In this context, where established technologies are merging with new ones, medicine is transitioning from discriminative models, which are algorithms designed to classify data, such as detecting fractures on radiographs, to generative systems capable of creating new content, including complex text, code, image, video, and synthetic data.[6,7]

Recent evaluations across orthopaedic subspecialties, including hip preservation by Slawaska-Eng et al,[8] rotator cuff pathology by Hand et al,[9] and shoulder arthroplasty by Lack et al,[10] revealed that while AI holds promise, its unguided application is fraught with variability in accuracy and reliability, necessitating a structured approach for scientific use.

For the academic orthopaedic surgeon, the pressure to "publish or perish" often collides with the increasing demands of clinical practice and patient care.[11] Within this constrained professional environment, the production of high-quality scholarly output becomes not only an expectation but an important logistical challenge.

The traditional process of conducting systematic reviews or writing original manuscripts is notoriously labor-intensive, often requiring months of data extraction and synthesis. Furthermore, for non-native English speakers, language barriers can pose major hurdles to publication in high-impact journals, creating an equity gap in scientific dissemination.[12]

This challenge is exacerbated by the fact that, while English is the native language of less than 10% of the global population and dominates nearly 75% of academic publications, this often requires non-native researchers to invest considerably more time and resources in manuscript preparation.[13] This linguistic imbalance not only increases the cognitive and logistical burden on researchers but may also influence study design, reporting clarity, and peer-review outcomes, thereby reinforcing structural inequities in scientific visibility and career advancement.[14]

Fortunately, the emergence of LLM-based systems, such as *ChatGPT*,[15] *Google Gemini*,[16] *Google NotebookLM*,[17] *Claude*,[18] *Perplexity*,[19] among others, offers a unique opportunity to democratize access to high-level scientific production. These tools can function as "research assistants," automating tedious tasks ranging from literature screening to linguistic polishing.[20] However, the adoption of these tools is still met with skepticism regarding accuracy and ethics. AI has been described as a paradigm-shifting advance, yet its irresponsible use may lead to "scientific pollution."[21]

Bridging this gap between potential and risk requires more than technological adoption alone. This article provides a practical framework for orthopaedic clinicians to utilize these technologies. To ensure proper and ethical use, study teams should ideally include or consult with members who have training in research methodology, data science, or applied AI, alongside domain expertise in orthopaedics. We delineate a workflow that moves from "black box" generation to grounded, verifiable research assistance, leveraging GenAI for manuscript drafting while strictly adhering to ethical guidelines on authorship and accountability.

METHODOLOGY

Literature Search Strategy

The literature addressing GenAI in the context of orthopaedic research and scientific writing is recent and rapidly evolving, with a marked increase in publications over the past few years. The semantic discovery tool *Consensus*[22] was used iteratively to explore terminology and refine the free-text keyword strategy before finalizing the PubMed[23] query (Appendix A, item 1.1). The final search was conducted in PubMed using the following free-text query: (ChatGPT OR "large language model" OR "generative artificial intelligence" OR GenAI) AND (orthop* OR musculoskeletal) AND (writing OR manuscript OR publication OR research).

Thus, the PubMed search was conducted on January 1, 2026, and encompassed publications available up to that date. Article screening and selection were performed independently by two reviewers, both authors of this manuscript, with discrepancies resolved by consensus. Given the narrative and methodological nature of this work, no formal exclusion criteria based on study design or outcomes were applied, as outlined in renowned checklists such as PRISMA[24]; instead, articles were included if they addressed GenAI in orthopaedic research, scientific writing, or editorial practice. Foundational methodological and policy-oriented sources not indexed under orthopaedic terms were systematically included through the snowballing method.[25,26]

Ultimately, this manuscript aims to provide a concise, narrative overview of current practice, rather than an exhaustive or systematic synthesis of the literature. Given the rapidly evolving nature of GenAI technologies and their recent application to orthopaedic research and scientific writing, the emphasis was placed on conceptual clarity, practical workflow design, and ethical considerations, rather than comprehensive literature enumeration.

Framework Development and Expert Review

The proposed framework was developed through an iterative, expert-informed process that included a critical appraisal of the literature and adherence to the authors' established editorial standards. This approach does not constitute a DELPHI process,[27] a formal consensus methodology, or an evidence-based guideline, but rather represents a narrative synthesis refined through discussions with two experts in GenAI, one a data scientist from an academic research institution and the other a technology professional from the private sector with expertise in applied AI systems, and practical experience in orthopaedic research and scientific writing.

DECODING THE TECHNOLOGY BEHIND GENERATIVE AI

Responsible use of GenAI in orthopaedic research depends on distinguishing between underlying model architectures, deployed tools, and their appropriate applications, as failure to do so directly contributes to unverified and erroneous scientific output.

Probabilistic Foundations of Text Generation

All GenAI tools are developed on probabilistic models trained on vast datasets to predict the next word in a sequence. These models are exceptional at generating coherent text, summarizing ideas, and refining grammar, however, they suffer from a critical flaw known as "hallucination."[28] It has been warned that LLMs can fabricate facts, data, and even non-existent citations with high confidence.[29] In addition, it was demonstrated that when asked to generate references[30], *ChatGPT* fabricated citations and provided inaccurate DOI[31] numbers for a substantial portion of them.

This limitation was starkly corroborated in a study on reverse shoulder arthroplasty, where *ChatGPT* scored zero on the JAMA benchmark criteria[32] for providing reliable citations[10]. Furthermore, the "native voice" of these models may be unsuitable for general communication without refinement. It was found that *ChatGPT*-generated content had a grade level of 14.7, equivalent to a university graduate level, which is considerably more complex than standard educational materials.[9] Therefore, relying solely on a standard LLM for factual data extraction or literature searching is dangerous and scientifically unsound.

Anchoring AI in Verifiable Evidence

Retrieval-Augmented Generation (RAG) is an established approach in GenAI that combines retrieval and generation.[33] It is a widely used paradigm for GenAI applications, particularly for knowledge-intensive tasks and question answering, as it augments the language model's inherent capabilities with external information, grounding outputs in authentic, traceable information.[34]

This data can come from three primary sources:

1. **Researcher-curated orthopaedic source content**, including peer-reviewed journal articles, clinical guidelines, consensus statements, and study protocols that provide domain-specific scientific context;
2. **Source-derived representations and data structures**, such as unstructured manuscript text, structured tables of extracted variables, semi-structured figures and charts, and vector embeddings generated from uploaded PDF documents to enable efficient semantic retrieval;
3. **External bibliographic and indexing resources**, including publicly available databases, journal metadata, and citation indexes used to contextualize findings and support literature discovery.

Together, these inputs enable RAG systems to generate outputs that are more accurate, transparent, and contextually grounded for orthopaedic scientific writing.

These enhancements help mitigate the hallucination phenomenon, but do not eliminate it completely. Importantly, as the size and scope of the supplied knowledge base grow, the potential for retrieval of tangential or contradictory information increases, which may itself introduce errors if not carefully managed. Furthermore, an incomplete, inappropriately constructed, or biased source corpus may lead to biased and misleading outputs, as the system can only synthesize from the data it is given. RAG connects the LLM to a specific, trusted knowledge base, such as a folder of PDF articles selected by the clinician.[35] When a user asks a question, the system first retrieves relevant text chunks from these specific documents and then uses the LLM to synthesize an answer based only on that retrieved data.

Recently, GenAI tools have incorporated this principle, and most notably, tools such as *Google NotebookLM* primarily ground their responses in external data. By grounding the AI's output in verifiable sources provided by the user, RAG considerably reduces, but does not eliminate, the risk of data fabrication, requiring immediate verification.[36]

AI-ENHANCED ORTHOPAEDIC RESEARCH WORKFLOW

A three-phase workflow is presented to assist the surgeon-scientist in efficiently producing high-quality manuscripts, as depicted in Figure 1. Table 1 summarizes the functional categories of AI-enabled systems involved at each phase, detailing their primary roles, typical applications, associated risks, and required verification strategies.

Phase 1: Literature Search and Curation

The first step is to identify relevant studies. Traditional scientific databases, such as PubMed/MeSH, Scopus, and Web of Science,[37,38] among others, should remain the backbone, while AI semantic discovery systems serve as high-yield complements for research discovery. This dual strategy enhances completeness and efficiency, ensuring that seminal papers and newly emerging studies are both captured early in the research process.

Notably, AI-powered discovery engines utilize "semantic search," which understands the intent and context of a query rather than just matching keywords. These tools typically index open-access articles and, depending on the platform, may also interface with major bibliographic databases such as PubMed, Scopus, and Web of Science, although coverage varies and may not include all subscription-restricted content. Unlike traditional systematic screening workflows, these platforms do not replicate the manual title-abstract-full-text screening pipeline; instead, they accelerate the discovery phase by surfacing semantically related papers and citation networks.[39] They should be used as complementary tools to identify connections between papers and execute the snowballing process of discovering valuable related documents.[40] In addition, these platforms can perform citation-network mapping, visualize research clusters, highlight author collaborations, and reveal emerging trends or gaps within a topic area, functions that accelerate topic familiarization and guide question refinement.

• **The tools:** Platforms such as *Elicit,*[41] *ResearchRabbit,*[42] or *Consensus.*

• **Application:** Tools like *Elicit* can parse semantic concepts to find papers that traditional search strings might miss.[43] *ResearchRabbit* creates visual citation networks, allowing clinicians to identify seminal documents and follow the "genealogy" of a research topic, ensuring that no major study is overlooked.[3] Nevertheless, general LLMs can be utilized to complement search strategies before querying the scientific databases; It has been noted that *ChatGPT* can identify relevant Medical Subject Headings (MeSH) terms more efficiently than the "MeSH on Demand" service generator from the National Library of Medicine,[44] thereby enhancing the categorization and searchability of the research[45].

The tools allow clinicians to employ natural language prompts such as:

- "What are the clinical outcomes of quadriceps tendon autograft vs. hamstrings for ACL reconstruction?"
- "How does robotic-assisted total knee arthroplasty compare with conventional techniques in terms of alignment accuracy?"
- "What are the long-term complication rates associated with 3D-printed spinal implants?"

Importantly, these prompts should not be treated as fixed queries but as part of an iterative, interactive process in which initial questions are progressively refined based on retrieved evidence, emerging knowledge gaps, and domain expertise.

This process is documented in the audit trail (Appendix A, item 1.1), which illustrates the employment of *Consensus* for the **Literature Search Strategy** section of this manuscript.

• Methodological Contributions

Beyond simplifying literature discovery, this phase provides tangible benefits that strengthen the rigor, transparency, and efficiency of the research process:

- Builds a curated library of high-quality, topic-specific PDFs that anchors the entire research workflow.
- Ensures traceability and reproducibility, allowing every extracted claim to be verified against primary sources.
- Enables visualization of citation networks and thematic clusters, helping identify influential papers and research gaps.
- Provides a verified evidence base for RAG and AI-assisted synthesis, grounding outputs in peer-reviewed literature.

Phase 2: Data Extraction and Synthesis

Once the curated library is built, the clinician must extract relevant data from the included studies.[46] AI tools can perform qualitative data extraction to identify conceptual relationships, including factors that influence surgical outcomes or trends across implant generations, thereby transforming isolated findings into structured, analyzable knowledge.[47] Beyond quantitative data, the extraction may include clinical variables, such as sample size, demographics, surgical technique, graft type, fixation method, and follow-up duration, methodological features, namely study design, level of evidence, inclusion/exclusion criteria, and statistical methods, and outcome metrics like functional scores, complication rates, imaging findings, and return-to-sport timelines.[25]

Traditionally, this phase has been a major bottleneck for research; however, using standard tools, such as *ChatGPT* or *Google Gemini*, for this purpose remains risky due to context-window limitations and the high risk of hallucination noted earlier.

• **The tools:** Platforms such as *Google NotebookLM*, *SciSpace*,[48] *Elicit*, or custom RAG-based tools.

• **Application:** Upload curated sources, importing the selected PDFs into the platform. This approach creates a "closed system" in which the AI's responses are grounded in and restricted to the information in your files.[49]

Structured Prompting

In **Phase 1**, highly structured prompting is not initially required, as well-formed natural language queries are generally sufficient to interrogate semantic search tools and retrieve relevant literature**.** In **Phases 2** and **3**, however, more highly structured prompts become necessary, as these stages require controlled information extraction, refined instructions, and constrained outputs. For this reason, it is essential to deliberately "engineer" a prompt using detailed, precise, and task-specific instructions that clearly define both the expected inputs and outputs.

Prompt Engineering refers to the deliberate crafting of inputs or instructions designed to optimize a language model's performance and output [50–52]. It ensures that GenAI systems produce results that are coherent, contextually relevant, and aligned with user intent. Prompt engineering is essential to ensure that AI-generated outputs meet rigorous standards of clarity, precision, and relevance.[52] A well-constructed prompt has a direct influence on the quality and usefulness of responses. Precise prompts yield more accurate, efficient, and trustworthy outcomes, whereas poorly designed prompts can lead to ambiguity or model hallucinations.[53]

An example of a structured natural language prompt for quantitative data extraction:

> *"**Task:***
> *Extract data from the uploaded orthopaedic studies.*
>
> ***Instruction:***
> *From each paper, extract and summarize the following variables:*
> *Author*
> *Year of publication*
> *Sample size*
> *Mean follow-up duration*
> *Graft type*
> *Lysholm score*
> *Failure rate*
>
> ***Output:***
> *Present the extracted information in a structured table with one column per variable listed above."*

This prompt illustrates a structured approach to data extraction in orthopaedic research. By clearly defining the task, specifying input variables, and constraining the output format, the researcher minimizes ambiguity and ensures the reproducibility of their results.

This process is documented in the audit trail (Appendix A, item 1.2), which illustrates the extraction of grounded qualitative data later used for the **Hallucinations and Accuracy** section of this manuscript.

Ensuring Reliability Through Human Oversight

RAG-based tools, in general, provide citations and footnotes linking the generated data point back to the specific location in the PDF. Nevertheless, the clinician must verify these data points. It has been demonstrated that, although AI-based extraction is efficient, human oversight remains necessary to capture nuance.[54] While no statistically significant difference in accuracy was found between *ChatGPT*-3.5 and 4.0 for questions on femoroacetabular impingement,[8] it has also been reported that approximately 30% of AI responses regarding shoulder arthroplasty were classified as "poor" by expert surgeons.[10] In conclusion, a RAG-based system mitigates hallucinations and inaccuracies by anchoring outputs to the verified uploaded text, without dispensing with human validation of the results. This oversight is crucial because AI performance remains inconsistent.

• Methodological Contributions

Beyond simplifying data extraction, this phase enhances methodological rigor and transparency in several key ways:

- Builds structured evidence tables directly from verified sources, reducing human error and increasing reproducibility.
- Enables rapid visualization of extracted quantitative or qualitative data, supporting meta-analysis and evidence synthesis.
- Strengthens traceability, linking each data point to its precise location within the source PDF.
- Provides a validated, AI-ready dataset that feeds seamlessly into manuscript drafting and analytical GenAI workflows.

Phase 3: Manuscript Drafting

With the extracted, literature-grounded data in hand, the clinician can use GenAI tools for the text-drafting process. It is vital to approach the GenAI role in this phase as a drafter and polisher, or even as a rigorous reviewer, rather than as the intellectual author.

• **The tools:** Platforms such as *ChatGPT*, *Google Gemini*, *Perplexity*, *Claude*, or custom RAG-based tools.

• **Application:**

The application of GenAI at this stage follows a sequential, evidence-aware workflow that prioritizes preparation of source materials, incremental drafting, and iterative human validation to ensure coherence, traceability, and alignment with established scholarly writing practices.

Organizing the content: Before feeding information into a GenAI tool, gather and systematically structure all relevant materials, including raw data generated in **Phase 2**, such as synthesized tables, and complementary sources, such as notes, reports, protocols, and reference documents. This preparatory step ensures coherence, traceability to original evidence, and minimizes the risk of factual distortion or contextual loss during automated text generation.[55]

Drafting approach: Rather than instructing a GenAI system to produce an entire manuscript in a single prompt, authors should adopt a section-by-section drafting strategy that mirrors the conventional scholarly writing process.[56] The IMRAD structure (Introduction, Methods, Results, Discussion) is a minimal standard format for any primary scientific report.[57]

It is essential that, before engaging GenAI assistance, the clinician first drafts the paragraphs for the intended section and defines its substantive content independently. Only after this initial author-driven step should the tool be used interactively to critique, refine, and enhance the text's clarity, structure, and linguistic quality.[58]

Additionally, the drafted paragraphs may be supplemented with complementary sources embedded or attached to the selected LLM, followed by a prompt requesting refinements and improvements.

This prompt example supports the refinement of the initial draft text:

> ***"Task:***
> *Refine the author-written draft for the Introduction section to improve academic clarity, coherence, and sentence-level quality while preserving the author's original meaning, argument, and evidentiary scope.*
>
> ***Input:***
>
> - *Attached file for contextualization: ResearchAim.docx*
> - *Draft text:*
>
> *[Insert draft text here]*
>
> ***Instructions:***
> *A draft text and supporting attachments were provided for the Introduction section of the manuscript.*
>
> *First, assess the draft paragraph for:*
>
> - *Clarity and logical flow*
> - *Sentence structure and conciseness*
> - *Consistency with formal academic medical writing*
> - *Factual support based only on the attached files*
>
> *Then, revise the paragraph accordingly, subject to the following constraints:*
>
> - *Base all assessments and revisions solely on the provided draft and attached files*
> - *Do not introduce new data, concepts, interpretations, references, or claims*
> - *Do not infer information that is not explicitly supported by the draft or the attached files*

• Perform a language-level and structure-level refinement only; preserve the author's intended meaning and argumentative position
• If any statement is unclear, overstated, or unsupported by the attached files, flag it explicitly rather than correcting it by invention
• Where changes are made, explain the reason for each change briefly and precisely

Output:

1. A revised version of the paragraph
2. A brief list of suggested changes, each with justification
3. A short note identifying any statements that require author verification"

This constrained prompt exemplifies how GenAI can be used as a structured editorial aid, guiding critique and refinement while preserving evidentiary boundaries and maintaining the primacy of author judgment.

It was found that AI-generated outlines are often logical and comprehensive, serving as a backbone structure for the manuscript.[55] Nevertheless, the process should always be validated and discussed by the authors and external reviewers.[59] This approach limits contextual overload, reduces factual distortion, and preserves authorial intent.

This process is documented in the audit trail (Appendix A, item 1.3), which illustrates the revision of the text later used for the **The Danger of Homogenization** section of this manuscript.

• **Combining GenAI and non-GenAI systems:** Even prior to any data analysis, GenAI systems may assist in outlining statistical strategies, clarifying variable relationships, or validating the coherence of the proposed analytical approach against the study objectives.[60]

Core research activities, such as data curation, statistical testing, and quantitative modeling, must remain grounded in conventional, validated software environments, including spreadsheets for data management and *SPSS*[61] or *R*[62] for statistical analysis. These tools ensure methodological rigor, reproducibility, and transparency, and GenAI does not displace their role.

After results are generated using conventional statistical tools, GenAI can be employed iteratively to help structure the narrative description of findings, test the internal logic of interpretations, and ensure that conclusions are appropriately aligned with the observed data, without altering or rederiving numerical outputs.[33]

Within this framework, GenAI tools are not used to generate results but rather to support research planning, analytical planning, and interpretive reasoning. Ultimately, the objective of integrating GenAI into the manuscript drafting process is not to replace established analytical or statistical tools but to support researchers in using them more effectively and in articulating their findings with greater clarity and coherence.

An example of a prompt that can support and critically assess the statistical approach in a manuscript, either before formal analysis planning or after the use of specialist statistical tools to support the analysis and conclusions:

*"***Task**:
Support the author in critically reviewing, refining, and validating the statistical analysis strategy for an orthopaedic surgery study.

Input:
Attached file: DataAnalysis.xlsx
Analysis and conclusions draft text:

[Your draft text here]

Instructions:
Use only the information explicitly provided in the draft text and attached file.
Assess whether the proposed statistical methods are appropriate for:

- *The study design*
- *The variables and outcomes*
- *The research questions*

Critically identify:

- *Methodological weaknesses*
- *Blind spots in the analysis strategy*
- *Unsupported analytical choices*
- *Missing statistical details that limit interpretability or validity*

Output:

1. Critical assessment of the statistical approach
2. Weaknesses, blind spots, inconsistencies, or missing information"

In addition, if the researcher prefers a more granular or interactive exploration of specific aspects of the study, complex prompts can be decomposed into smaller, thematically focused blocks, such as study design, variable selection, statistical assumptions, or outcome definitions, allowing for component-level examination while maintaining strict control over scope and evidentiary grounding.

• **Refinement for Non-Native Speakers:** For clinicians whose first language is not English, GenAI acts as a sophisticated editor. Del Giglio and Costa[12] emphasize that AI can rephrase complex technical descriptions into fluent, academic English, thereby leveling the playing field for international publication. Consider adding context instructions to the prompt, focusing on the output format, to help tools ensure alignment with scholarly conventions.

This prompt example will help to refine the text into British English:

*“**Task:***
Refine the provided paragraph for language quality and academic style.

Input:

[Your draft text here]

Instructions:

• Revise the text to improve grammar, clarity, and flow using British English spelling and conventions, appropriate for a peer-reviewed orthopaedic surgery journal
• Do not add new information, references, or interpretations
• Do not alter numerical values, technical terminology, or scientific meaning
• If any sentence is ambiguous, preserve the original meaning rather than clarifying through inference

Output:
Return the revised paragraph in polished academic British English”

As a final safeguard, all AI-refined text should undergo review and validation by a domain expert and, whenever possible, a native English-speaking reviewer, who retains full responsibility for the accuracy, nuance, and scientific intent of the final manuscript.

• Methodological Contributions
Beyond supporting manuscript drafting, this phase contributes to research quality, clarity, and accountability in several key ways:

- Maintains authorial control and accountability, as all substantive content, interpretations, and conclusions are drafted and validated by the clinician, with AI limited to structured critique and refinement.
- Improves coherence and clarity of the manuscript, supporting alignment between study objectives, methods, results, and conclusions through section-by-section drafting.
- Supports interpretive rigor without altering results, enabling logical consistency checks and clearer narrative description while preserving all numerical outputs generated by conventional statistical tools.
- Reduces linguistic and editorial barriers, particularly for non-native English speakers, while ensuring final review by domain experts and retention of scientific intent.

RISKS AND ETHICAL CONSIDERATIONS

While efficient, this workflow introduces specific risks that must be managed to maintain scientific integrity.

Hallucinations and Accuracy

Despite the use of RAG-based systems, errors may still occur. Hallucinations may occur, leading to fabricated or distorted information being presented as fact if users are insufficiently vigilant.[29] The fluency of AI-generated text further complicates the task of detecting it. Blinded reviewers failed to distinguish between AI-generated and human-written abstracts in 32% of cases.[45] It has similarly been demonstrated that early AI-generated reviews, while time-efficient, contained substantial reference inaccuracies, reaching up to 70% in non-RAG workflows.[55]

LLMs are unable to independently assess the evidentiary hierarchy or the quality of sources without human oversight. The absence of reliable citations in *ChatGPT*-generated orthopaedic content has been highlighted, underscoring the need for systematic verification.[10] Moreover, the opaque and proprietary nature of model training data limits transparency regarding source weighting and framing, introducing additional uncertainty.[63] The rapid scalability of text generation has also facilitated unethical practices, including "paper mills." It has been estimated that approximately 3% of medical publications in 2022 exhibited features consistent with those of mass-produced, fraudulent manuscripts.[64]

Action: To mitigate hallucinations and factual inaccuracies, researchers must implement systematic human verification at every stage of AI-assisted work. All extracted data, citations, and synthesized statements should be cross-checked against the original source PDFs or primary literature before inclusion. Maintaining a structured audit trail, recording prompts, model outputs, and verification steps, ensures transparency and reproducibility.[65,66]

Minimum audit trail elements, recommendation: For each AI-assisted task, the audit trail should document (1) the AI tool and model used; (2) the date of interaction; (3) the purpose of the interaction; (4) the manuscript location of the intervention; (5) the final version of the prompt; (6) the source corpus supplied to the model; (7) the corresponding human verification step, including confirmation, correction, or rejection with reference to the original source. This documentation may be maintained as a structured log or appendix and should be retained for internal review, peer review if requested, and post-publication accountability. As a practical example, the audit trail for the present manuscript is provided as Appendix A in the supplementary file.

Authorship and Accountability

There is a broad consensus among editorial bodies that GenAI tools do not meet authorship criteria, as authorship requires accountability, approval of the final manuscript, and responsibility for the integrity of the work, capacities that AI systems do not possess, as outlined by Yoo[67], the International Committee of Medical Journal Editors (ICMJE),[68] and Silva et al.[35]

Overreliance on AI-generated critique or structuring may also promote cognitive offloading, diminishing active engagement in data interpretation and scholarly reasoning.[69] In addition, AI outputs are not static; responses may vary across model versions or updates, limiting reproducibility unless AI use is transparently documented.[58]

Action: Human authors must take full responsibility for every word in the manuscript. A classification system has been proposed for reporting AI involvement, urging authors to disclose the specific tool used and the scope of its application.[70] The major scientific journal publishers, including Sage,[71] Emerald,[72] Wiley,[73] Elsevier,[74] IEEE,[75], and Taylor & Francis,[76] have established formal policies governing the responsible use of GenAI in scholarly writing, consistently emphasizing author accountability, transparency, and disclosure.

Bias and Data Privacy

GenAI models may reproduce or amplify biases embedded in their training data. The underrepresentation of specific populations in training corpora can lead to biased framing of disease prevalence, risk factors, and outcomes.[77] Beyond training data limitations, generative models may also amplify dominant narratives during synthesis, potentially underrepresenting minority findings or negative results unless actively counterbalanced by domain-expert review.[63,78]

Equally important are risks related to data privacy and confidentiality. Information entered into publicly accessible AI platforms may be retained or incorporated into future model training, potentially leading to unauthorized disclosure of data.[79] In the research context, this raises important concerns regarding compliance with data protection regulations, including HIPAA[80] and the General Data Protection Regulation (GDPR).[81] Additionally, the terms of service of many commercial AI platforms may permit the secondary use or retention of submitted content, creating ambiguity regarding data ownership and intellectual property rights.[82,83]

These concerns extend beyond authorship to the peer-review process. Reviewers who upload submitted manuscripts to GenAI tools for summarization or critique may inadvertently breach copyright protections and violate the confidentiality owed to authors.[64]

Action: Personally identifiable information, patient-level data, and unpublished manuscripts must never be entered into open or non–institutionally governed AI platforms.[84,85] Researchers should use only secure, institutionally approved systems that comply with data protection standards such as HIPAA and GDPR. At present, no comprehensive, publicly maintained registry of GDPR- and HIPAA-compliant AI tools exists for each step of the research workflow described herein. The development of such a resource, potentially maintained by professional societies such as ISAKOS, would be a valuable contribution to the field. All AI-assisted workflows must include safeguards for data anonymization and confidentiality, with clear documentation of where and how data are processed.[4,86] In addition, researchers should critically assess AI outputs for potential bias, ensuring that interpretations are reviewed and balanced by human domain experts to prevent the reinforcement of inequities or misrepresentation of findings.

The Danger of Homogenization

The widespread, unregulated use of GenAI in medical writing may progressively homogenize scientific discourse, eroding originality and critical thinking.[87] When multiple authors rely on the same models and prompting strategies, manuscripts risk converging toward a standardized narrative style, potentially diminishing the diversity of perspectives that underpin scientific advancement.[21,88]

Empirical evidence supports this concern. It has been demonstrated that AI-generated orthopaedic content consistently converges toward a uniform, postgraduate-level reading complexity, which may inadvertently narrow accessibility and contribute to stylistic uniformity across publications.[9] Such convergence risks diluting nuanced argumentation and obscuring novel interpretations derived from clinical expertise.

Action: To mitigate this risk, surgeon-scientists must retain full ownership of data interpretation, argument development, and narrative framing. GenAI should function exclusively as a supportive editorial tool, enhancing clarity and structure without dictating content or tone[13,89]. Authors should intentionally preserve their disciplinary voice, critical reasoning, and stylistic individuality to maintain diversity in scientific expression.[58] Upholding this balance ensures that AI strengthens scholarship without diluting originality or the interpretive depth derived from clinical expertise.

CONCLUSION AND FUTURE PERSPECTIVES

Orthopaedics is entering an era in which GenAI is increasingly integrated into the research workflow. When applied as a *cognitive exoskeleton*, these tools can significantly reduce the administrative and logistical burden of scientific work, enabling clinicians to devote more attention to data interpretation, methodological rigor, and clinical relevance.

By adopting the structured framework proposed in this article, leveraging semantic search systems for discovery, retrieval-augmented generation for verifiable data extraction, and LLMs for controlled drafting and refinement, orthopaedic surgeons can improve both the efficiency and quality of their scholarly output. Importantly, this integration must remain firmly human-centered. GenAI can assist the research process, but it cannot substitute for domain expertise, critical reasoning, or ethical judgment.

Scientific writing in orthopaedics should therefore not be shaped by automation replacing the surgeon-scientist, but by clinicians responsibly harnessing AI to augment their intellectual work. When used within clearly defined boundaries, supported by rigorous verification and transparent disclosure, GenAI has the potential to enhance academic productivity while preserving scientific integrity and authorship accountability.

Limitations

This study has several limitations that should be considered when interpreting its findings and recommendations.

First, this work is conceptual and methodological in nature rather than empirical. The proposed framework represents an expert-informed synthesis of existing literature, editorial policies, and current practice, rather than a consensus statement or prospectively validated intervention. Consequently, although the workflow is grounded in widely accepted best practices, its impact on research quality, efficiency, or reproducibility has not been quantitatively evaluated.

Second, the manuscript is based on a focused narrative literature search, which, while transparently reported, does not aim to provide an exhaustive or systematic synthesis of all available evidence. As such, selection bias cannot be entirely ruled out, and the framework should be interpreted as a state-of-the-practice overview rather than a comprehensive evidence map.

Third, the framework is intentionally limited to text-based research workflows and manuscript preparation. Its applicability to other domains, including imaging analysis, surgical video interpretation, biomechanical modeling, or real-time clinical decision support, was beyond the scope of this work and would require distinct validation strategies and ethical safeguards.

Fourth, the proposed approach assumes a baseline level of methodological literacy and digital competence among users. Surgeons with limited experience in research methods, statistics, or AI-enabled tools may derive less benefit and may be more vulnerable to misuse or overreliance without appropriate training and institutional governance.

Fifth, recommendations are necessarily influenced by the rapidly evolving nature of GenAI technologies. Model capabilities, accessibility, licensing conditions, and publisher policies are subject to change, which may limit the long-term generalizability of specific tool-based examples described in this manuscript.

Sixth, although ethical risks such as hallucinations, bias, privacy, and authorship are addressed in detail, risk mitigation remains fundamentally dependent on humans. No framework can fully prevent inappropriate use, insufficient verification, or ethical lapses in the absence of active human oversight, institutional support, and adherence to journal standards.

Finally, this work does not address cost, infrastructure, or access disparities associated with advanced AI tools. Differences in institutional resources, data governance frameworks, and regional regulatory environments may constrain equitable adoption, particularly in low-resource settings.

Despite these limitations, this manuscript aims to provide a practical and ethically grounded framework that reflects current realities in orthopaedic research and scientific writing.

Future Directions

Looking ahead, the emergence of agentic AI systems, capable of autonomously performing multi-step tasks by chaining prompts, retrieving information, and executing actions across connected tools, may further extend the role of AI in orthopaedic research workflows. These systems extend traditional GenAI by adding reasoning and task orchestration, thereby enabling partial workflow automation under human oversight.[34,90]

The responsible integration of agentic AI will require robust governance frameworks to ensure transparency, controllability, and unambiguous attribution of responsibility. Accordingly, future research should prioritize the empirical validation of AI-assisted workflows, the assessment of real-world performance of agentic systems, and the establishment of best practices for their ethical deployment within academic research environments.

In parallel, efforts should focus on developing standardized evaluation criteria and reporting conventions for AI-assisted research in orthopaedics. Establishing shared benchmarks for documentation, verification, disclosure, and accountability would enhance comparability across studies and support consistent editorial and peer-review assessment. Such standardization may facilitate the controlled integration of agentic AI tools into academic practice while preserving methodological rigor and alignment with journal and institutional policies.

Finally, authors are encouraged to use, adapt, and prospectively test the proposed audit trail in their own AI-assisted research workflows, thereby contributing to its refinement, contextual validation, and broader methodological standardization.

-

Declaration of Generative AI and AI-assisted technologies in the writing process

Statement: *During the preparation of this work, the authors used Consensus (powered by GPT-5.0) for complementary literature exploration, Google NotebookLM (powered by Gemini 2.5) to ground and support claims in the text and to create Figure 1, and ChatGPT (powered by GPT-5.2) to assess and critique the draft. After using these tools, the authors reviewed and edited the content as needed and took full responsibility for the publication's content.*

REFERENCES

1. Auricchio F, Marconi S. 3D printing: clinical applications in orthopaedics and traumatology. *EFORT Open Reviews*. 2016;1(5):121-127. doi:10.1302/2058-5241.1.000012

2. Lang JE, Mannava S, Floyd AJ, et al. Robotic systems in orthopaedic surgery. *The Journal of Bone and Joint Surgery British volume*. 2011;93-B(10):1296-1299. doi:10.1302/0301-620X.93B10.27418

3. Singh RP, Karuparti A, Maeda J. *Generative AI for Cloud Solutions: Architect Modern AI LLMs in Secure, Scalable, and Ethical Cloud Environments*. 1st ed. Packt Publishing Limited; 2024.

4. Islam R. *Generative AI, Cybersecurity, and Ethics*. John Wiley & Sons, Inc.; 2025.

5. Salminen J, Jung S gyo, Almerekhi H, Cambria E, Jansen B. How Can Natural Language Processing and Generative AI Address Grand Challenges of Quantitative User Personas? In: Degen H, Ntoa S, Moallem A, eds. *HCI International 2023 – Late Breaking Papers*. Springer Nature Switzerland; 2023:211-231.

6. Oettl FC, Pruneski JA, Zsidai B, et al. Is orthopaedics entering the age of generative AI?—A narrative review of current applications, challenges, and future directions. *Knee surg sports traumatol arthrosc*. Published online October 27, 2025:ksa.70145. doi:10.1002/ksa.70145

7. Ibrahim M, Khalil YA, Amirrajab S, et al. Generative AI for synthetic data across multiple medical modalities: A systematic review of recent developments and

challenges. *Computers in Biology and Medicine*. 2025;189:109834. doi:10.1016/j.compbiomed.2025.109834

8. Slawaska-Eng D, Bourgeault-Gagnon Y, Cohen D, Pauyo T, Belzile EL, Ayeni OR. ChatGPT-3.5 and -4 provide mostly accurate information when answering patients' questions relating to femoroacetabular impingement syndrome and arthroscopic hip surgery. *Journal of ISAKOS*. 2025;10:100376. doi:10.1016/j.jisako.2024.100376

9. Hand C, Bohn C, Tannir S, et al. American Academy of Orthopaedic Surgeons OrthoInfo provides more readable information regarding rotator cuff injury than ChatGPT. *Journal of ISAKOS*. 2025;12:100841. doi:10.1016/j.jisako.2025.100841

10. Lack BT, Mouhawasse E, Childers JT, et al. Can ChatGPT answer patient questions regarding reverse shoulder arthroplasty? *Journal of ISAKOS*. 2024;9(6):100323. doi:10.1016/j.jisako.2024.100323

11. Mavrogenis AF, Scarlat MM. Artificial intelligence publications: synthetic data, patients, and papers. *International Orthopaedics*. 2023;47(6):1395-1396. doi:10.1007/s00264-023-05830-w

12. Giglio AD, Costa MUPD. The use of artificial intelligence to improve the scientific writing of non-native english speakers. *Rev Assoc Med Bras*. 2023;69(9):e20230560. doi:10.1590/1806-9282.20230560

13. Kayaalp ME, Ollivier M, Winkler PW, et al. Embrace responsible ChatGPT usage to overcome language barriers in academic writing. *Knee surg sports traumatol arthrosc*. 2024;32(1):5-9. doi:10.1002/ksa.12014

14. Bryant A, Charmaz K, eds. *The SAGE Handbook of Current Developments in Grounded Theory*. SAGE Publications, Inc.; 2019. doi:10.4135/9781526436061

15. OpenAI. ChatGPT. ChatGPT. 2022. Accessed April 7, 2024. https://chat.openai.com

16. Google. Google Gemini. Google Gemini. 2023. Accessed December 15, 2025. https://gemini.google.com

17. Google. Google NotebookLM | Note Taking & Research Assistant Powered by AI. Google NotebookLM. 2023. Accessed February 6, 2025. https://notebooklm.google.com/

18. Anthropic. Claude. 2023. Accessed December 15, 2025. https://claude.ai/new

19. Perplexity. Perplexity. Perplexity AI. 2022. Accessed December 15, 2025. https://www.perplexity.ai/

20. Khalifa M, Albadawy M. Using artificial intelligence in academic writing and research: An essential productivity tool. *Computer Methods and Programs in Biomedicine Update*. 2024;5:100145. doi:10.1016/j.cmpbup.2024.100145

21. Golan R, Reddy R, Muthigi A, Ramasamy R. Artificial intelligence in academic writing: a paradigm-shifting technological advance. *Nat Rev Urol*. 2023;20(6):327-328. doi:10.1038/s41585-023-00746-x

22. Consensus. Consensus: AI for Research. Consensus. 2022. Accessed December 16, 2025. https://consensus.app/

23. National Library of Medicine. PubMed. PubMed. 1996. Accessed January 1, 2026. https://pubmed.ncbi.nlm.nih.gov/

24. Page MJ, Moher D, Bossuyt PM, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ*. Published online March 29, 2021:n160. doi:10.1136/bmj.n160

25. Aromataris E, Lockwood C, Porritt K, Pilla B, Jordan Z, eds. *JBI Manual for Evidence Synthesis*. JBI Global; 2024. doi:10.46658/JBIMES-24-01

26. Badampudi D, Wohlin C, Petersen K. Experiences from using snowballing and database searches in systematic literature studies. In: *Proceedings of the 19th International Conference on Evaluation and Assessment in Software Engineering*. ACM; 2015:1-10. doi:10.1145/2745802.2745818

27. Dalkey N, Helmer O. An Experimental Application of the DELPHI Method to the Use of Experts. *Management Science*. 1963;9(3):458-467. doi:10.1287/mnsc.9.3.458

28. Ahaley SS, Pandey A, Juneja SK, Gupta TS, Vijayakumar S. ChatGPT in medical writing: A game-changer or a gimmick? *Perspectives in Clinical Research*. 2024;15(4):165-171. doi:10.4103/picr.picr_167_23

29. Alkaissi H, McFarlane SI. Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. *Cureus*. Published online February 19, 2023. doi:10.7759/cureus.35179

30. Athaluri SA, Manthena SV, Kesapragada VSRKM, Yarlagadda V, Dave T, Duddumpudi RTS. Exploring the Boundaries of Reality: Investigating the Phenomenon of Artificial Intelligence Hallucination in Scientific Writing Through ChatGPT References. *Cureus*. Published online April 11, 2023. doi:10.7759/cureus.37432

31. Paskin N. Digital Object Identifier (DOI®) System. In: *Encyclopedia of Library and Information Sciences*. 4th ed. CRC Press, Taylor & Francis Group; 2017:1325-1331. https://www.taylorfrancis.com/chapters/edit/10.1081/E-ELIS4-133/digital-object-identifier-doi®-system-norman-paskin

32. Silberg WM. Assessing, Controlling, and Assuring the Quality of Medical Information on the Internet: Caveant Lector et Viewor—Let the Reader and Viewer Beware. *JAMA*. 1997;277(15):1244. doi:10.1001/jama.1997.03540390074039

33. Alkhalaf M, Yu P, Yin M, Deng C. Applying generative AI with retrieval augmented generation to summarize and extract key clinical information from electronic

health records. *Journal of Biomedical Informatics*. 2024;156:104662. doi:10.1016/j.jbi.2024.104662

34. Huang K, Huang G, Dawson A, Wu D. GenAI Application Level Security. In: Huang K, Wang Y, Goertzel B, Li Y, Wright S, Ponnapalli J, eds. *Generative AI Security: Theories and Practices*. Future of Business and Finance. Springer Nature Switzerland; 2024:199-237. doi:10.1007/978-3-031-54252-7_7

35. Silva JCMC, Gouveia RP, Zielinski KMC, et al. AI-Assisted Tools for Scientific Review Writing: Opportunities and Cautions. *ACS Appl Mater Interfaces*. 2025;17(34):47795-47805. doi:10.1021/acsami.5c08837

36. Lewis P, Perez E, Piktus A, et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *arXiv*. Preprint posted online 2020. doi:10.48550/ARXIV.2005.11401

37. Pranckutė R. Web of Science (WoS) and Scopus: The Titans of Bibliographic Information in Today's Academic World. *Publications*. 2021;9(1):1. doi:10.3390/publications9010012

38. AlRyalat SAS, Malkawi LW, Momani SM. Comparing Bibliometric Analysis Using PubMed, Scopus, and Web of Science Databases. *JoVE (Journal of Visualized Experiments)*. 2019;(152):e58494. doi:10.3791/58494

39. Dhakal A, Paudel K, Sigdel S. An Artificial Intelligence Driven Semantic Similarity-Based Pipeline for Rapid Literature Review. *arXiv*. Preprint posted online 2025. doi:10.48550/ARXIV.2509.15292

40. Jalali S, Wohlin C. Systematic literature studies: database searches vs. backward snowballing. In: *Proceedings of the ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*. ACM; 2012:29-38. doi:10.1145/2372251.2372257

41. Elicit, inc. Elicit: AI for scientific research. Elicit. 2023. Accessed December 16, 2025. https://elicit.com/

42. ResearchRabbit. ResearchRabbit: AI Tool for Smarter, Faster Literature Reviews. ResearchRabbit. 2021. Accessed December 16, 2025. https://www.researchrabbit.ai/

43. Kung J. Elicit (product review). *J Can Health Libr Assoc*. 2023;44(1). doi:10.29173/jchla29657

44. U.S. National Library of Medicine. National Library of Medicine - National Institutes of Health. National Library of Medicine. 1993. Accessed December 16, 2025. https://www.nlm.nih.gov/

45. Dahmen J, Kayaalp ME, Ollivier M, et al. Artificial intelligence bot ChatGPT in medical research: the potential game changer as a double-edged sword. *Knee surg sports traumatol arthrosc*. 2023;31(4):1187-1189. doi:10.1007/s00167-023-07355-6

46. Bolaños F, Salatino A, Osborne F, Motta E. Artificial intelligence for literature reviews: opportunities and challenges. *Artif Intell Rev*. 2024;57(10):259. doi:10.1007/s10462-024-10902-3

47. Cooper H, Hedges LV, Valentine JC, eds. *The Handbook of Research Synthesis and Meta-Analysis*. 3rd ed. Russell Sage Foundation; 2019. doi:10.7758/9781610448864

48. PubGenius Inc. AI Chat for scientific PDFs | SciSpace. SciSpace. 2022. doi:10.3402/ijch.v74.29482

49. Scherbakov D, Hubig N, Jansari V, Bakumenko A, Lenert LA. The emergence of large language models as tools in literature reviews: a large language model-assisted systematic review. *Journal of the American Medical Informatics Association*. 2025;32(6):1071-1086. doi:10.1093/jamia/ocaf063

50. Evelyn L. *Making ChatGPT Work for You: Getting the Most Out of Generative AI as a Non-Techie*. Apress; 2025. doi:10.1007/979-8-8688-1445-7

51. OpenAI. Prompt engineering | OpenAI API. OpenAI Platform. 2023. Accessed January 2, 2026. https://platform.openai.com

52. Koch D, Kohne A, Brechbühler N. Fundamentals of Prompt Engineering. In: *Prompt Engineering in the Enterprise – An Introduction: Competitive Advantages through Generative AI and Large Language Models*. Springer Fachmedien Wiesbaden; 2025:21-33. doi:10.1007/978-3-658-49218-2_2

53. Qadir J. Engineering Education in the Era of ChatGPT: Promise and Pitfalls of Generative AI for Education. In: *2023 IEEE Global Engineering Education Conference (EDUCON)*. IEEE; 2023:1-9. doi:10.1109/EDUCON54358.2023.10125121

54. Wang S, Scells H, Zhuang S, Potthast M, Koopman B, Zuccon G. Zero-shot Generative Large Language Models for Systematic Review Screening Automation. *arXiv*. Preprint posted online 2024. doi:10.48550/ARXIV.2401.06320

55. Kacena MA, Plotkin LI, Fehrenbacher JC. The Use of Artificial Intelligence in Writing Scientific Review Articles. *Curr Osteoporos Rep*. 2024;22(1):115-121. doi:10.1007/s11914-023-00852-0

56. Dwivedi YK, Kshetri N, Hughes L, et al. Opinion Paper: “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*. 2023;71:102642. doi:10.1016/j.ijinfomgt.2023.102642

57. Booth A, Sutton A, Clowes M, Martyn-St James M. *Systematic Approaches to a Successful Literature Review*. Third edition. SAGE Publications, Inc.; 2022.

58. European Commission. Living guidelines on the responsible use of generative AI in research | Research and innovation. Published online 2024. Accessed March 21,

2024. https://research-and-innovation.ec.europa.eu/document/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en

59. Zielinski C, Winker MA, Aggarwal R, et al. Chatbots, generative AI, and scholarly manuscripts: WAME recommendations on chatbots and generative artificial intelligence in relation to scholarly publications. *Colomb Med*. 2023;54(3):e1015868. doi:10.25100/cm.v54i3.5868

60. Alto V. *Practical Generative AI with ChatGPT: Unleash Your Prompt Engineering Potential with OpenAI Technologies for Productivity and Creativity*. 2nd ed. Packt Publishing Limited; 2025.

61. Morgan GA, Barrett KC, Leech NL, Gloeckner GW. *IBM SPSS for Introductory Statistics: Use and Interpretation*. 6th ed. Routledge, Taylor & Francis Group; 2019. doi:10.4324/9780429287657

62. Schumacker RE. *Learning Statistics Using R*. SAGE Publications, Inc.; 2015. doi:10.4135/9781506300160

63. Buick A. Copyright and AI training data—transparency to the rescue? *Journal of Intellectual Property Law and Practice*. 2025;20(3):182-192. doi:10.1093/jiplp/jpae102

64. Wascher DC, Ollivier M. Large Language Models in Orthopaedic Publications: The Good, the Bad and the Ugly. *Am J Sports Med*. 2024;52(9):2193-2195. doi:10.1177/03635465241265692

65. Sweenor DE, Mulkers Y. *Generative AI Business Applications: An Executive Guide with Real-Life Examples and Case Studies*. TinyTechMedia LLC; 2024.

66. Review HB. *Generative AI: The Insights You Need from Harvard Business Review*. Harvard Business Review Press; 2024.

67. Yoo JH. Defining the Boundaries of AI Use in Scientific Writing: A Comparative Review of Editorial Policies. *J Korean Med Sci*. 2025;40(23):e187. doi:10.3346/jkms.2025.40.e187

68. Editors IC of MJ. Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. Updated April 2025. Published online 2025. https://www.icmje.org/icmje-recommendations.pdf

69. Singh S, Kumar R, Maharshi V, et al. Harnessing Artificial Intelligence for Advancing Medical Manuscript Composition: Applications and Ethical Considerations. *Cureus*. Published online October 17, 2024. doi:10.7759/cureus.71744

70. Hryciw BN, Seely AJE, Kyeremanteng K. Guiding principles and proposed classification system for the responsible adoption of artificial intelligence in scientific writing in medicine. *Front Artif Intell*. 2023;6:1283353. doi:10.3389/frai.2023.1283353

71. Sage. Artificial intelligence policy. Sage Journals. 2023. Accessed December 30, 2025. https://www.sagepub.com/journals/publication-ethics-policies/artificial-intelligence-policy

72. Emerald. Emerald Publishing's stance on AI tools and authorship | Emerald Publishing. Emerald Publishing. 2023. Accessed December 30, 2025. https://www.emeraldgrouppublishing.com/news-and-press-releases/emerald-publishings-stance-ai-tools-and-authorship

73. Wiley. Notice to Authors. Wiley Online Library. 2023. Accessed December 30, 2025. https://onlinelibrary.wiley.com/page/journal/18673899/homepage/notice-to-authors

74. Elsevier. Publishing ethics | Elsevier policy. Elsevier Policies. 2023. Accessed December 30, 2025. https://www.elsevier.com/about/policies-and-standards/publishing-ethics

75. IEEE. Submission and Peer Review Policies. IEEE Author Center Journals. 2023. Accessed December 30, 2025. https://journals.ieeeauthorcenter.ieee.org/become-an-ieee-journal-author/publishing-ethics/guidelines-and-policies/submission-and-peer-review-policies/

76. Robinson M. Taylor & Francis Clarifies the Responsible use of AI Tools in Academic Content Creation. Taylor & Francis Newsroom. February 17, 2023. Accessed December 30, 2025. https://newsroom.taylorandfrancisgroup.com/taylor-francis-clarifies-the-responsible-use-of-ai-tools-in-academic-content-creation/

77. Crossnohere NL, Elsaid M, Paskett J, Bose-Brill S, Bridges JFP. Guidelines for Artificial Intelligence in Medicine: Literature Review and Content Analysis of Frameworks. *J Med Internet Res*. 2022;24(8):e36823. doi:10.2196/36823

78. OpenAI. Is ChatGPT biased? OpenAI Help Center. 2023. Accessed January 7, 2026. https://help.openai.com/en/articles/8313359-is-chatgpt-biased

79. Kim SJ. Research ethics and issues regarding the use of ChatGPT-like artificial intelligence platforms by authors and reviewers: a narrative review. *Science Editing*. 2024;11(2):96-106. doi:10.6087/kcse.343

80. The United States Department of Health and Human Services. HIPAA Home. The United States Department of Health and Human Services. June 9, 2021. Accessed December 24, 2025. https://www.hhs.gov/hipaa/index.html

81. European Commission. *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA Relevance)*. 2016. Accessed July 17, 2025. https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng

82. Musiol M. *Generative AI: Navigating the Course to the Artificial General Intelligence Future*. 1st ed. John Wiley & Sons, Inc.; 2024.

83. Reddy S. Generative AI in healthcare: an implementation science informed translational path on application, integration and governance. *Implementation Sci*. 2024;19(1):27. doi:10.1186/s13012-024-01357-9

84. Alavi M, Leidner D, Mousavi R. Knowledge Management Perspective of Generative Artificial Intelligence (GenAI). *SSRN Journal*. Published online 2024. doi:10.2139/ssrn.4782875

85. Gupta M, Akiri C, Aryal K, Parker E, Praharaj L. From ChatGPT to ThreatGPT: Impact of Generative AI in Cybersecurity and Privacy. *IEEE Access*. 2023;11:80218-80245. doi:10.1109/access.2023.3300381

86. Fui-Hoon Nah F, Zheng R, Cai J, Siau K, Chen L. Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. *Journal of Information Technology Case and Application Research*. 2023;25(3):277-304. doi:10.1080/15228053.2023.2233814

87. Ramoni D, Sgura C, Liberale L, Montecucco F, Ioannidis JPA, Carbone F. Artificial intelligence in scientific medical writing: Legitimate and deceptive uses and ethical concerns. *European Journal of Internal Medicine*. 2024;127:31-35. doi:10.1016/j.ejim.2024.07.012

88. Brown O, Davison RM, Decker S, et al. Theory-Driven Perspectives on Generative Artificial Intelligence in Business and Management. *British J of Management*. 2024;35(1):3-23. doi:10.1111/1467-8551.12788

89. Stokel-Walker C, Van Noorden R. What ChatGPT and generative AI mean for science. *Nature*. 2023;614(7947):214-216. doi:10.1038/d41586-023-00340-6

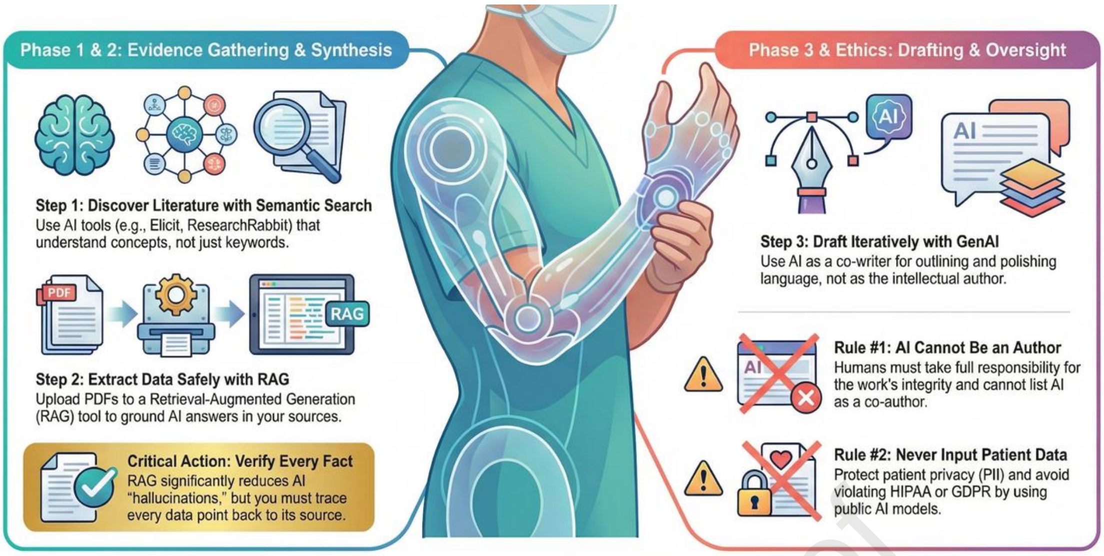
90. Thomas R, Zikopoulos P, Soule K. *AI Value Creators: Beyond the Generative AI User Mindset*. O'Reilly Media, Inc.; 2025.

FIGURE LEGENDS

Figure 1. The AI research assistant: A Surgeon's Framework for Scientific Writing.

TABLES

Table 1. AI Tools in Orthopaedics Scientific Writing: Purposes, Risks, and Verification Strategies



AI system category (by epistemic role)	Primary function in the research workflow	Best use cases in orthopaedic writing	Key risks	Verification/mitigation strategies
Authoritative bibliographic retrieval systems (PubMed, Scopus, Web of Science)	Indexed literature retrieval based on controlled vocabularies, editorial curation, and citation standards	Establishing the core evidence corpus; ensuring completeness, traceability, reproducibility, and index-quality coverage	Omission of relevant studies due to terminology mismatch, indexing delays, or database-specific coverage limits	Use as the foundational backbone of all reviews; complement with concept-based discovery tools to expand recall while preserving bibliographic rigor and transparency
AI-assisted discovery systems (semantic, retrieval-first; e.g., Elicit, ResearchRabbit, Consensus)	Concept-based literature discovery and citation network exploration without autonomous content generation	Early topic exploration; identification of seminal studies; forward and backward citation snowballing; refinement of research questions and scope	Algorithmic opacity; unstable or non-transparent relevance ranking; incomplete or non-reproducible coverage	Cross-check results against conventional databases; manually screen and document inclusion decisions; avoid treating algorithmic ranking as a proxy for study quality or level of evidence
Evidence-grounded generative systems (closed-corpus RAG workflows; e.g.,	Source-constrained extraction and synthesis are strictly limited to user-supplied documents	Structured extraction of cohorts, outcomes, follow-up, complications,	Context fragmentation, selective or partial extraction, and anchoring	Require explicit in-text anchors or citations to source PDFs; verify each extracted data point against original documents; maintain an

Google NotebookLM)		and methods; construction of verifiable evidence tables anchored to primary sources	bias toward retrievable text segments	auditable extraction and verification log
General-purpose language models (open-generation LLMs, optionally RAG-enabled; e.g., ChatGPT, Google Gemini, Claude, Perplexity)	Text structuring, drafting support, critique, and linguistic refinement based on author-provided inputs	IMRAD outlining; section-by-section drafting; logical flow and coherence checks; editorial refinement and language support for non-native English authors	Hallucinated facts or references; lack of awareness of evidentiary hierarchy; overconfident or homogenized narrative tone	Restrict use to author-supplied content; prohibit LLM-only factual claims; independently verify all references (DOI/PMID); retain full human control and accountability over interpretation and conclusions

Declarations of interest:

The authors declare that they have no conflict of interest related to this work.