# Using AI to boost scoping reviews. Exploring AI deployment in obstetrics and gynaecology as an exemplar

Isabella M. Venter[1[0000-0001-6554-8024]], Rénette J. Blignaut[1[0000-0002-3150-6624]] and Karen V. Renaud[2,3 [0000-0002-7187-6531]]

[1] University of the Western Cape, Cape Town, South Africa
[2] University of Strathclyde, Glasgow, Scotland
[3] University of South Africa, Pretoria, South Africa
rblignaut@uwc.ac.za; karen.renaud@strath.ac.uk

**Abstract.** This paper explores how artificial intelligence can be harnessed to boost and supercharge scoping reviews. To demonstrate its power (comprehensiveness), its efficiency (time effectiveness) and affordability (cost effectiveness), we explore the use of AI in obstetrics and gynaecology as a vehicle to test our new hybrid review process. We carried out: (1) a traditional scoping review, and (2) a traditional scoping review boosted by artificial intelligence tools. We provide an overview of the extensive and varied literature related to our topic of interest: the use of AI in obstetrics and gynaecology. We compare and contrast the outcomes of the different options. We conclude that the use of AI tools can enrich and extend the scope of scoping reviews (using a so-called "Hybrid Review"), but only if the prompts are carefully and thoughtfully crafted.

**Keywords:** artificial intelligence, large language models, chatbots, obstetrics, gynaecology, diagnosis, prognosis, ethics.

## 1 Introduction

Natural language processing (NLP)—a field in artificial intelligence (AI)—is being used by Large Language Models (LLMs) across numerous fields, including healthcare. LLMs are trained on massive datasets and use billions of parameters to perform tasks such as answering questions, translating text, and completing sentences [1]. Models such as the Generative Pre-trained Transformer (GPT) [2] and Gemini [3] have demonstrated the ability to produce coherent and contextually appropriate text. Since the introduction of the chatbot ChatGPT by OpenAI on 21 November 2022, many similar bots have been developed. Poe.com, for example, is a website that lists several of these bots: official bots, reasoning bots, search bots, image generating bots, text analysis bots, etc. These AI-powered bots generate responses dynamically, often using large language models or external databases for context.

However, these AI tools' use in research and practice is constrained by significant challenges, such as their tendency to generate inaccurate or misleading information, often referred to as "hallucination" [4]. Furthermore, LLM's are time constrained—i.e.,

it is limited by the training data which is at best a few months old [5] and thus the information they provide is often not current.

Retrieval-Augmented Generation (RAG) is a method that enhances the output of a large language model by integrating information from an authoritative external knowledge base, beyond the model's original training data, before generating a response [6]. RAG builds on the powerful capabilities of LLMs by enabling them to access domain-specific information without requiring model retraining. This makes RAG a cost-effective solution for improving the relevance, accuracy, and applicability of LLM outputs across a variety of contexts (see **Fig. 1**). For example, in healthcare, RAG can be utilised to handle requests for relevant patient records or research papers and generate personalised treatment plans. For example, it will be able to analyse a patient's medical history and combine this with recent research insights to recommend tailored treatment solutions just prior to the initiation of care [5].

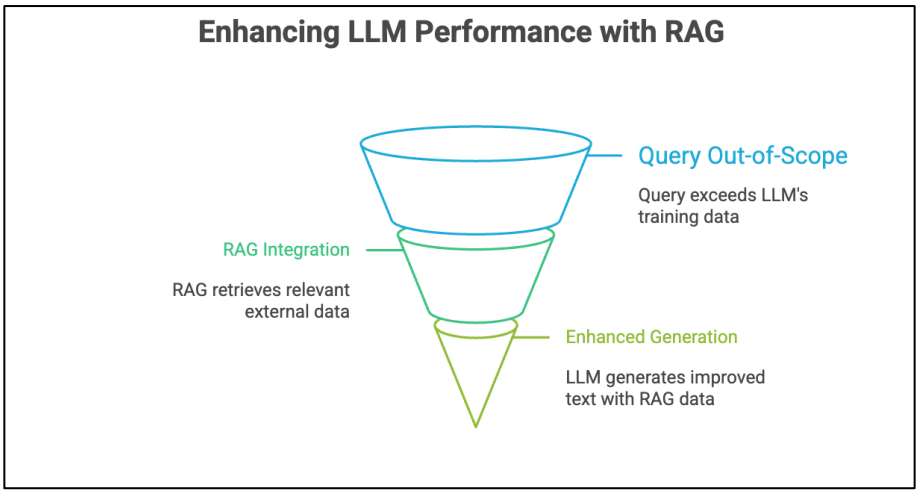


**Fig. 1**: Example of how RAG is used to add recent or more specific information (generated by Napkin.ai)

Although AI has been adopted in academia and in education, many faculty members, especially in the health sciences, are still unsure of its benefits or how to advise its ethical deployment. As reported in a recent study "*There was broad consensus among faculty members that students should receive training on how to use AI chatbots*" [7] (p. 5). Haleem *et al.* [8] argue that the use of AI is going to change all domains of medicine. According to Weidener and Ficher [9] although many health science students engage with AI technology, particularly AI chat applications, only a small percentage have undergone formal AI training. This indicates a significant gap in current health science curricula. It underscores the need to update health science education to include AI and AI ethics, ensuring that future healthcare professionals are well-prepared to address the challenges and opportunities that AI presents in healthcare [9].

Scoping reviews are incredibly powerful in terms of giving researchers a snapshot of what researchers have found related to any topic [12]. Traditional scoping reviews have strong power in terms of delivering a comprehensive set of papers. It is also affordable in terms of the tool being free to anyone who wants to use it. However, it is extremely time consuming. AI-powered reviews, on the other hand, are extremely efficient—identifying papers in a matter of minutes. However, to be able to harness their full power, researchers must pay for the premier version. As such, these tools are not universally affordable to many researchers.

**Fig. 2** shows how the different review tools perform on three measures: power (comprehensiveness), affordability (cost effectiveness) and efficiency (time effectiveness). Traditional scoping reviews are affordable, given that most universities provide access to researchers. If carried out rigorously, it has maximum power to find all relevant publications. It is, however, extremely time consuming, so does not exhibit efficiency. AI-Powered reviews are very efficient, delivering their outcomes in a matter of minutes. To get the full benefit of their power, a paid subscription is usually required, which reduces their affordability. They do not have a full measure of power, given that they do not reveal the same comprehensive coverage as traditional reviews. A hybrid review, on the other hand, can deliver the best of both worlds.

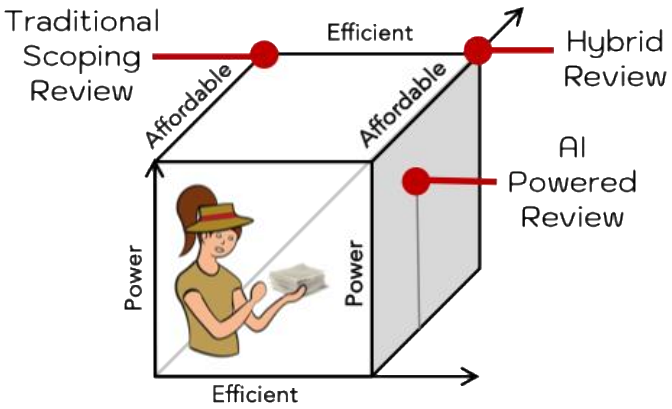


**Fig. 2**: Characteristics of scoping review tools (drawn by the authors)

This study examined how artificial intelligence can be used to discover its deployment in healthcare with specific reference to obstetrics and gynaecology as proof of concept. The research question being posed is: *How can AI help streamline the scoping review process of identifying current AI applications in all domains?*

This question can be translated into two sub-questions:

**RQ1.** *To what extent can hybrid tools (traditional searches augmented with AI) boost scoping reviews, in terms of power, efficiency and affordability?*

***RQ2.*** *How has AI been used in obstetrics and gynaecology?*

A scoping review and a review using AI tools were used to address the first research question. The results were compared, and it was found that AI tools simplify the search and identification of appropriate papers considerably. The only limitation being the AI tools' affordability. Furthermore, using AI to do the content analysis it was found that AI is being used mostly for diagnostic imaging and predictive analytics within the field of obstetrics and gynaecology. The main contribution of this paper is to suggest a hybrid review process, harnessing both traditional and AI-powered searches. We tested our new process by searching for papers in one specific branch of medicine, but it will apply equally to other searches.

## 2 Materials & Methods

Pragmatism, as a research approach, focuses on understanding the world and addressing its challenges by combining both subjective and objective methods. Pragmatist researchers adopt a flexible and practical approach to finding effective solutions for the phenomena being studied and acknowledge the value of incorporating diverse perspectives and worldviews to tackle research problems [10].

In higher education, especially in the context of AI, pragmatism offers a framework for educators and institutions to navigate technology-enhanced learning environments. This perspective aligns with John Dewey's educational theories, which stress that knowledge should be rooted in real-world experiences and that education should prepare individuals for practical life [11].

### 2.1 Methodology

A pragmatic approach—which emphasises the practical implications of research findings and integrating various methods to address the research question—was used for this research. The methods employed were a scoping review using PRISMA and a review using AI tools—SciSpace, Elite Beta and Consensus. These AI tools were randomly chosen from AI tools that offer free options. Finally, the results were compared (see **Fig. 3**) and content analysis was used to address the research questions.

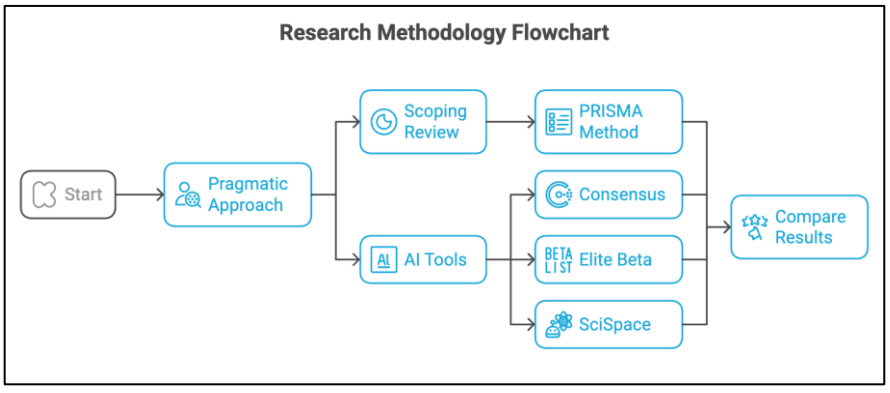


**Fig. 3** Research approach (defined by the researchers and generated by Napkin.ai)

## 2.2 Methods

The scoping review as method, is suitable since "…*scoping reviews do not aim to produce a critically appraised and synthesised result/answer to a particular question, …it rather aims to provide an overview or map of the evidence*." [12] (p. 3).

**Scoping review using PRISMA**. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) was primarily intended for systematic reviews of research evaluating the effects of health interventions but has since been adopted by many other fields of study (see **Fig. 4**).

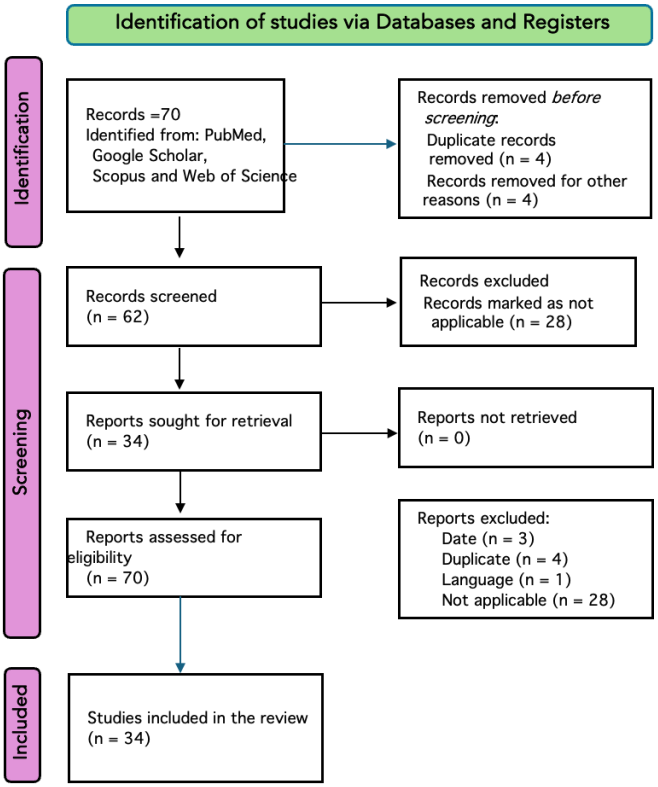


**Fig. 4**: Prisma flowchart of articles identified, screened and included in the study

As can be seen from the figure, it has three main steps for reviewing literature: identification of articles, screening of the articles and deciding which articles to include in the review.

Using the search word/phrases: AI and obstetrics, AI and gynaecology, AI and obstetrics and gynaecology, AI and obstetrics and gynaecology and education, 70 papers that could be downloaded in full were identified by the researchers using scholarly databases. Twenty-five articles were found using PubMed, 18 using Google Scholar, 18

using Scopus and 9 using Web of Science. All of these were published in the period 2024–2025.

No automation tools—such as Rayyan—were used to generate the PRISMA flowchart, all the literature were evaluated by two researchers. Of the 70 articles only 34 were retained for review. Four articles were removed since they were duplicates, another 4 were removed because of incorrect publication dates and one being written in German. Twenty-eight articles were removed because they dealt with medical education but not specifically with the training of obstetrics and gynaecology medical students or registrars. This search and identification of relevant papers was carried out over several days and took approximately 10 hours altogether.

**Review using AI tools.** It was decided to implement some freely available AI tools such as SciSpace, Elicit Beta, and Consensus to identify relevant articles. More AI tools are available such as: Semantic Scholar, Research Rabbit, Connected, and LitSence. etc. these were however not used in this investigation.

SciSpace, was interrogated—on the 20th of April 2025—to identify journal articles for the first research question, namely *"How is AI being used in obstetrics and gynaecology in terms of diagnosis, its challenges and benefits?"* SciSpace has three levels of searches, namely Standard, High Quality or Deep Review. Deep Review was chosen and 1050 papers were identified of which it suggested 243 were relevant to the research question. It furthermore wrote a short summary of the "top" 20 papers—it seems as if it refers to the most recent papers identified—in terms of practical applications in clinical settings, theoretical advancements and research and development: It could also be instructed to write a summary of the top 10 or 5 papers. When the same request was made on the 30th June 2025, only the standard review could be accessed without payment. The search identified 100 relevant journal papers with a summary of the first 5 papers without payment. For some articles the PDF could be requested, for others it could be downloaded directly. This search was used for the comparison.

Elicit Beta allowed the user to use some filters—for dates and quality of publications—to indicate what the search should include. When it was instructed (on the 20th of April 2025) to search for papers using the first research question it rephrased the research question to: "*What are the primary ethical, technical, and implementation challenges of integrating AI technologies in obstetric and gynaecological diagnostic practices?"* The free version allowed the consideration of 50 papers. However, according to the tool not all these papers' full texts were checked—only the abstract was checked for some of these identified papers. In its report Elicit Beta indicated that it searched across over 126 million academic papers from the Semantic Scholar corpus and retrieved the 50 papers most relevant to the query. On the 2nd of June 2025 when the research was repeated for the question "*What are the current applications of artificial intelligence in obstetrics and gynaecology for diagnosis and treatment?"* it identified 50 papers which according to it, is most relevant to the research question. The second search was used for the comparison.

When the same question was posed to Consensus, it provided 10 papers related to the research question as well as a summary of the content of the papers. Filters could be used and more papers—than the ten summarised—could be loaded. Information

about the papers such as the study population, study count, methods, outcomes and results could be retrieved. Consensus also allows for the download of a CSV file with a summary of the papers. The summary of all the information about the 10 papers was useful and was used for the comparison (see Fig. 5).

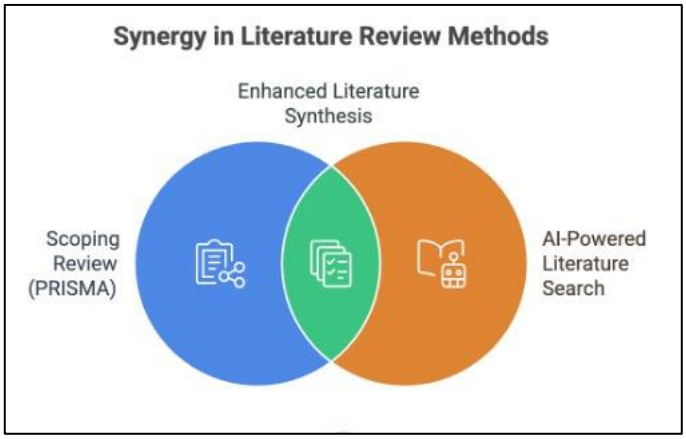


**Fig. 5**: Comparing the different methods (generated by Napkin.AI)

**Comparison.** The process using the AI tools was efficient —much more so than the traditional scoping review process. For the articles found, all the tools—SciScape, Elite Beta and Consensus—provided links which could be used to access the articles. That is, to confirm that the AI tools had not "*hallucinated*" and that these references and articles actually existed.

## 3 Results

It was decided to see which articles identified with the PRISMA method were also identified by the AI tools. As can be seen from Table 1, nine of the 34 articles identified using the PRISMA method, was also mentioned by one or more of the AI tools. Six was mentioned by Consensus, three by Elite Beta and 5 by SciSpace. Four articles were mentioned by three of the approaches. Only one article was not identified during the PRISMA search but was mentioned by all three of the AI tools.

**Table 1**: Comparing the PRISMA results with the results when using AI tools

| | *Title of the Article* | *DOI* | *PRISMA* | *Con-sensus* | *Elite Beta* | *Sci-Space* |
|---|---|---|---|---|---|---|
| 1 | Elbiss et al. Artificial intelligence in gynecologic and obstetric emergen-cies (2025) | 10.1186/s122 45-025-00820-8 | x | x | | |
| 2 | Saida, et al. Artificial Intelligence in Obstetric and Gyne-cological MR Imaging (2024) | 10.2463/ mrms.rev. 2024-0077 | x | x | x | |
| 3 | Patel et al. Artificial Intelligence in Obstetrics and Gynecology: Trans-forming Care and Outcomes (2024) | 10.7759/cu-reus.64725 | x | x | | x |

| | *Title of the Article* | *DOI* | *PRISMA* | *Consensus* | *Elite Beta* | *SciSpace* |
|---|---|---|---|---|---|---|
| 4 | Sone et al. Clinical Prospects for Artificial Intelligence in Obstetrics and Gynecology (2024) | 10.31662/jmaj.2024-0197 | x | x | | |
| 5 | Changhez et al. Evaluating the Efficacy and Accuracy of AI Assisted Diagnostic Techniques in Endometrial Carcinoma: A Systematic Review (2024) | 10.7759/cureus.60973 | x | | | x |
| 6 | Moro et al. Role of artificial intelligence applied to ultrasound in gynecology oncology: A systematic review (2024) | 10.1002/ijc.35092 | x | | x | x |
| 7 | Brandãoet al. Revolutionizing Women's Health: A Comprehensive Review of Artificial Intelligence Advancements in Gynecology (2024) | 10.3390/jcm13041061 | x | x | | |
| 8 | Gumilar et al. The promise and challenges of Artificial Intelligence-Large Language Models (AI-LLMs) in obstetric and gynecology (2024) | 10.20473/mog.v32i22024.128-135 | x | | | x |
| 9 | Mahmoudiandehkordi et al. Transforming Gynecology with Artificial Intelligence: Advances in Clinical Practice (2024) | 10.54756/ijsar.2024.25 | x | | | x |
| 10 | Chaurasia et al. Use of artificial intelligence in obstetric and gynaecological diagnostics: a protocol for a systematic review and meta-analysis (2024) | 10.1136/bmjopen-2023-082287 | | x | x | x |

**Content analysis.** Using ChatPDF, these ten papers were summarised. The summary of all ten papers where then used to prompt ChatGPT, Grok 3 and Deepseek, to extract the broad themes of the research. The researchers read through the summarised ten papers and identified specific themes.

As can be seen from **Table 2**, two cross-cutting themes were identified by all three AI tools and the researchers. These were: AI in diagnostic imaging and predictive analysis.

**Table 2:** Broad themes of the research

| *Theme* | *Grok 3* | *Deepseek* | *ChatGPT* | *Researchers* |
|---|---|---|---|---|
| AI in diagnostic imaging | x | x | x | x |
| Predictive analytics | x | x | x | x |
| Cancer detection & staging | x | x | | |
| Personalised medicine | x | | x | x |
| Ethical and practical considerations | x | | x | x |
| Reproductive medicine | | x | | |
| Emergency care | | x | | |
| Treatment planning | | | x | x |

| *Theme* | *Grok 3* | *Deepseek* | *ChatGPT* | *Researchers* |
|---|---|---|---|---|
| Challenges such as data quality, lack of standardisation, interpretability issues, and limited clinical validation | | | x | x |
| Human oversight | | | | x |

Personalised medicine and ethical considerations were mentioned by two of these tools and the researchers. Treatment planning and challenges such as data quality, lack of standardisation, interpretability issues, and limited clinical validation was mentioned by ChatGPT and the researchers. Other themes mentioned were AI's use for reproductive medicine, emergency care and the need for human oversight.

## 4 Discussion

This research considered how AI (using AI tools) can help streamline scoping reviews to identify current AI applications in all domains (see **Fig. 6**).

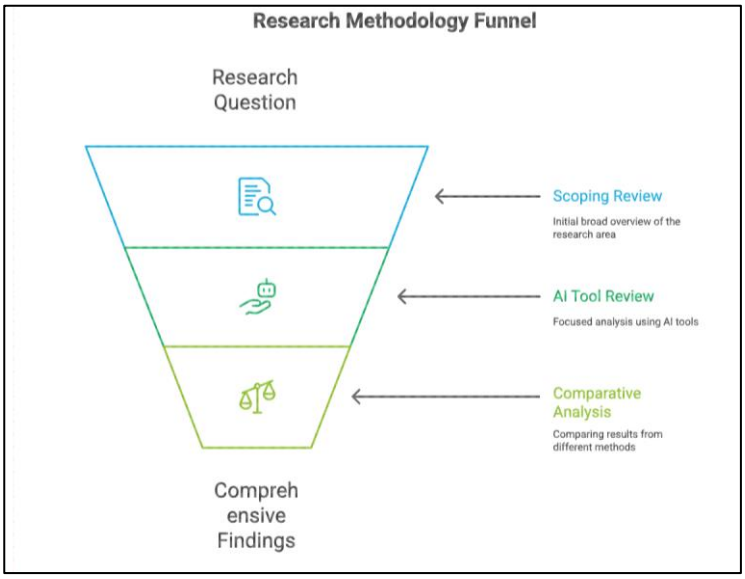


**Fig. 6**: Results using the research methodology funnel

### 4.1 Hybrid tools

The scoping review, using the PRISMA method, took approximately 10 hours and produced 34 articles that could be analysed. Whereas the review, using AI Tools, produced 160 papers in a matter of minutes providing—in most cases—a summary of the first 10 papers.

When all these papers were compared—those identified by both the traditional PRISMA method and using AI tools—it was found that nine papers were identified by at least one of the AI tools and PRISMA. To answer the first research question:

**RQ1**. *To what extent can hybrid tools (traditional searches augmented with AI) boost scoping reviews, in terms of power, efficiency and affordability?*

It became clear that using a hybrid approach can boost scoping reviews in terms of power—many relevant articles can be found in a short period, efficient in that ten of the papers were identified by most and affordable in that none of the tools used required a subscription.

Our conclusion is that AI tools can enrich and extend the scope of scoping reviews, but only if the prompts are carefully and thoughtfully crafted.

### 4.2 AI use in obstetrics and gynaecology

Analysing the ten papers that were identified, the second research question was addressed:

**RQ2**. *How has AI been used in obstetrics and gynaecology?*

It was found that AI could indeed assist with personalised medicine, ethical considerations, treatment planning but that there are challenges such as data quality, lack of standardisation, interpretability issues, and limited clinical validation. AI were mostly used in reproductive medicine, for cancer treatment and in emergency care. However, all of these require human oversight.

The question that may be asked when using AI for scoping reviews is: Are these identified articles representative of the research field globally? Furthermore, in terms of prompt engineering researchers need to be aware that when using an AI tool it does not remove bias from underlying training data and also that slight changes in prompts can yield different outputs, therefore it is important to test multiple prompt versions for reliability.

### 4.3 Future work

We suggest the following avenues for future research:

1. AI services usually only provide the first 10 results free of charge. The consequent lack of access to paid versions has a negative impact on comprehensiveness. Investigations into the impacts of different AI-tool pricing strategies would service to refine our hybrid review process.
2. In fleshing out the hybrid review approach, it would be good to explore the inclusion of an “expert in the loop” to help to decide on which tools to use and which papers to include in the final corpus. This would have an impact on power but is likely to impact time effectiveness.

## Acknowledgements

The first author sadly passed away a week before the paper's deadline. Since she had spearheaded this work, and had worked on the paper the day before she passed, her family wanted us to complete and submit the paper to honour her. The second and third authors express their appreciation for many years of valued collaborations with Isabella, for her friendship and her tremendous sense of fun, which made all interactions such a great pleasure. We are pleased to be able to present this paper as her final contribution to the research literature.

**List of Papers from Review**

1. Elbiss, H. M., Fikri M. A-Z.: Artificial intelligence in gynecologic and obstetric emergencies. International Journal of Emergency Medicine 18(1), Article 20 (2025). 10.1186/s12245-025-00820-8
2. Saida, T., Wenchao G., Sodai H., Toshitaka I., Masafumi S., Taishi A., Yuta N., Ayumi S., Toyomi S., Takahito N.: Artificial Intelligence in Obstetric and Gynecological MR Imaging. Magnetic Resonance in Medical Sciences 24(3), 354-365 (2025). 10.2463/mrms.rev.2024-0077
3. Patel, D. J., Chaudhari, K., Acharya, N., Shrivastava, D., Muneeba, S.: Artificial intelligence in obstetrics and gynecology: transforming care and outcomes. Cureus 16(7), e64725 (2024). 10.7759/cureus.64725
4. Sone, K., Taguchi, A., Miyamoto, Y., Uchino-Mori, M., Iriyama, T., Hirota, Y., Osuga, Y.: Clinical Prospects for Artificial Intelligence in Obstetrics and Gynecology. JMA Journal 8(1), 113-120 (2025) 10.31662/jmaj.2024-0197
5. Changhez, J., James, S., Jamala, F., Khan, S., Khan, M. Z., Gul, S., Zainab, I.: Evaluating the efficacy and accuracy of AI-assisted diagnostic techniques in endometrial carcinoma: A systematic review. Cureus 16(5), e60973 (2024). 10.7759/cureus.60973
6. Moro, F., Ciancia, M., Zace, D., Vagni, M., Tran, H. E., Giudice, M. T., Zoccoli, Mascilini, S. G. F., Ciccarone, F., Boldrini, L., D'Antonio, F., Scambia, G., Testa, A. C.: Role of artificial intelligence applied to ultrasound in gynecology oncology: A systematic review. International journal of cancer 155(10), 1832-1845 (2024). 10.1002/ijc.35092
7. Brandão, M., Mendes, F., Martins, M., Cardoso, P., Macedo, G., Mascarenhas, T., Saraiva, M. M.: Revolutionizing women's health: A comprehensive review of artificial intelligence advancements in gynecology. Journal of Clinical Medicine 13(4), 1061 (2024). 10.3390/jcm13041061
8. Gumilar, K. E., Tan, M.: The promise and challenges of Artificial Intelligence-Large Language Models (AI-LLMs) in obstetrics and gynecology. Journal Of Obstetrics & Gynecology Science 136(2), 128-135 (2024). 10.20473/mog.v32i22024.128-135
9. Mahmoudiandehkordi, S., Yeganegi, M. Transforming Gynecology with Artificial Intelligence: Advances in Clinical Practice. International Journal of Scientific and Applied Research (IJSAR), 4(9), 104-112 (2024). 10.54756/ijsar.2024.25

10. Chaurasia, A., Curry, G., Zhao, Y., Dawoodbhoy, F., Green, J., Vaninetti, M., Shah, N., Greer, O.: Use of artificial intelligence in obstetric and gynaecological diagnostics: a protocol for a systematic review and meta-analysis. BMJ Open 14(5), e082287 (2024). 10.1136/bmjopen-2023-082287

## References

[1] Floridi, L., Chiriatti, M.: GPT-3: Its Nature, Scope, Limits, and Consequences. Minds & Machines, 30, 681–694. (2020). 10.1007/s11023-020-09548-1
[2] OpenAI. (2023). GPT-4 Technical Report. Retrieved from: https://arxiv.org/abs/2303.08774.
[3] Google. (2023, December 6). Gemini. Retrieved from: Gemini.google.com: https://gemini.google.com
[4] Huang, L., W. Yu, W. Ma, W. Zhong, Z. Feng, H. Wang, Q. Chen, W. Peng, X. Feng, B. Qin, T. Liu.: A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. ACM Transactions on Information Systems 43(2), 1-55. (2025). 10.1145/3703155
[5] Amugongo, L. M., Mascheroni, P., Brooks, S., Doering, S., Seidel, J.: Retrieval augmented generation for large language models in healthcare: A systematic review. PLOS Digital Health 4(6), e0000877 (2025). 10.1371/journal.pdig.0000877
[6] Fan, W., Ding, Y., Ning, L., Wang, S., Li, H., Yin, D., T-S Chua, Li, Q. A survey on rag meeting LLMs: Towards retrieval-augmented large language models. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 6491-6501. Barcelona, Spain: ACM. (2024). 10.1145/3637528.3671470
[7] Saleh, Z. T., R. A. Elshatarat, M. Alharbi, B. N. Alhumaidi, M. S. Al-Za'areer, R. A. Jarrad, T. F. Al Niarat, W. T. Almagharbeh, K. M. Al-Sayaghi, D. El Sayed Fadila. Exploring faculty perceptions and concerns regarding artificial intelligence Chatbots in nursing education: potential benefits and limitations. BMC Nursing, 24(440), 1-13 (2025). 10.1186/s12912-025-03082-0
[8] Haleem, A., Javaid, M. & Khan, I.H. Current status and applications of Artificial Intelligence (AI) in medical field: An overview. Current Medicine Research and Practice 9(6), 231-237 (2019). 10.1016/j.cmrp.2019.11.005
[9] Weidener, L., & Fischer, M.: Artificial intelligence in medicine: cross-sectional study among medical students on application, education, and ethical aspects. JMIR medical education, 10(1), e51247. (2024). 10.2196/51247
[10] Elgeddawy, M., & Abouraia, M.: Pragmatism as a research paradigm. European Conference on Research Methodology for Business and Management Studies (pp. 71-74). Academic Conferences International Limited. (2024)
[11] Lin, F., Chiew Hwa, P., & Ku Wing, C.: The Influence of Dewey's Pragmatism Educational Thought on Chinese General Education: A Historical Summary. Malaysian Journal of Social Sciences and Humanities, 8(12), p. e002589 (2023). 10.47405/mjssh.v8i12.2589
[12] Munn, Z. P. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Medical Research Methodology, 18(143), 1-7. (2018). 10.1186/s12874-018-0611-x