

Utilizing Retrieval-Augmented Large Language Models for Pregnancy Nutrition Advice

Taranum Bano¹, Jagadeesh Vadapalli¹, Bishwa Karki¹, Melissa K Thoene²,
Matt VanOrmer², Ann L Anderson Berry², and Chun-Hua Tsai¹

¹ College of Information Science & Technology, University of Nebraska Omaha ,
Omaha, 68182, NE, USA

² College of Medicine, University of Nebraska Medical Center,
Omaha, 68188, NE, USA

Abstract. The importance of nutrition during pregnancy cannot be overstated, as it profoundly impacts maternal and fetal health outcomes. Optimal fetal growth and development are contingent upon adequate nutrition throughout gestation, which in turn requires that expectant mothers possess a high level of nutritional literacy. This latter factor may serve as a valuable predictor of pregnancy outcomes. This paper seeks to leverage the capabilities of a retrieval-augmented large language model to provide personalized prenatal nutrition guidance. We employed Meta’s LLAMA 2 model and integrated an expert-curated dataset of nutrition information. Our evaluation, conducted using ChatGPT-based metrics, revealed that while the augmented model did not yield significant improvements in overall response quality, it could generate more thoughtful and specific responses easily comprehensible to users. We conclude by discussing the challenges encountered and lessons learned from our investigation.

Keywords: Chatbot, Large Language Models, RAG, Health

1 Introduction

Nutrition is paramount during pregnancy due to its profound impact on maternal health and fetal development. The dietary choices made by expectant mothers significantly influence the growth and well-being of the fetus, as well as the overall health status of the mother [19]. Ensuring sufficient intake of essential nutrients like proteins, carbohydrates, fats, vitamins, and minerals is crucial for promoting fetal organ development, skeletal growth, and normal physiological function. Insufficient consumption of these vital nutrients can lead to developmental delays, restricted perinatal outcomes, and reduced newborn weight [15]. Of note, some women may consume more calories than recommended but experience malnutrition from a nutrient deplete diet of ultraprocessed foods [3]. Furthermore, maternal nutritional deficiencies may continue to affect the child’s growth and health even after birth. Studies have linked maternal malnutrition during pregnancy to various chronic illnesses such as obesity, diabetes, cardiovascular disease, and impaired cognitive development in offspring [15]. Adequate

nutrition not only mitigates the risk of adverse pregnancy outcomes but also contributes to long-term health trajectories for both mother and child. Thus, optimizing dietary intake during pregnancy is imperative to ensure optimal health outcomes and lay a foundation for lifelong well-being [15]. Despite the critical role of nutrition during pregnancy, expectant mothers face significant challenges in accessing reliable advice. The vast amount of conflicting and non-evidence-based information available, particularly online, often leaves them overwhelmed and uncertain about which sources to trust [14]. Additionally, the structure of prenatal care—with its infrequent appointments that need to cover a range of health concerns—may not allow for in-depth discussions on nutrition, thus impeding the absorption of necessary dietary information. Implementing supplementary support systems specifically designed to address pregnancy nutrition queries is therefore crucial. Such systems would provide a consistent and reliable framework for guidance, helping to alleviate the burden of navigating complex dietary recommendations on one’s own. Moreover, enhancing health literacy, which involves understanding and interacting effectively with the healthcare system, is essential for empowering individuals to make informed decisions about their health [18, 5]. This highlights the necessity for enhanced accessibility to accurate and comprehensible health information, which is crucial for effective healthcare engagement.

The potential of Large Language Models (LLMs) as a tool for providing personalized nutrition advice to pregnant patients has been explored, with models such as ChatGPT demonstrating promising results [20]. However, concerns have been raised regarding the accuracy and reliability of LLM-generated responses, which can be random, mistaken, or misleading [2]. To address this limitation, we propose combining LLMs with retrieval-augmented generation (RAG) techniques, which has shown promise in optimizing language model performance [10]. Integrating LLMs with RAG makes it possible to deliver users contextually relevant health information. Furthermore, by leveraging these advanced AI capabilities and expert-approved data embedding, we argue that high-quality, personalized health content can be generated, tailored to individual needs and preferences. The success of this design has the potential to bridge the gap in health literacy, empowering individuals to make informed decisions about their health and well-being [7].

This paper investigates whether LLMs enhanced with RAG can improve the quality of responses provided in pregnancy nutrition advice. To address this research question, we leveraged Meta’s LLAMA 2 models and integrated an expert-approved, self-collected dataset on pregnancy nutrition. We employed the state-of-the-art ChatGPT 4 model to evaluate the performance of our augmented model, which has been demonstrated to surpass human worker label quality [6]. Our evaluation revealed that the RAG-enhanced LLM may not significantly enhance the general information quality in the responses; however, the responses generated by this model were found to be more concise and easier to comprehend, potentially benefiting individuals with limited health literacy. This finding provides valuable insights into the efficacy of utilizing LLMs in nutrition

advising and the impact of incorporating self-collected data into this specific use case in health care. We subsequently discuss the implications, lessons learned, and conclusions drawn from our findings.

2 Related Works

2.1 Nutrition Literacy and Advice in Pregnancy

Research highlights the crucial role of nutrition during pregnancy, yet many women in the US struggle to meet recommended standards due to challenges like obesity and inadequate dietary intake [19]. Adolescent pregnant women, in particular, exhibit less healthy dietary patterns, emphasizing the need for enhanced support to optimize health outcomes for this vulnerable population [12]. Adequate intake of essential nutrients such as proteins, carbohydrates, fats, vitamins, and minerals is crucial for promoting fetal organ development, skeletal growth, and normal physiological functioning [15]. Moreover, maternal nutrition has long-term implications, influencing the risk of chronic diseases in offspring [9]. Additionally, a mother’s diet during pregnancy is linked to her mental well-being, with certain nutrients playing a role in promoting positive mental health [21]. Despite its importance, access to nutrition counseling remains limited, with systemic barriers, socioeconomic disparities, and high rates of unplanned pregnancies contributing to variations in access [8].

The importance of nutrition literacy, or the ability to understand available nutrition information, is highlighted by [16]. The study emphasizes that pregnant women are more likely to seek and follow nutrition advice as their nutrition literacy increases. Nutrition literacy among pregnant women is a critical aspect of maternal and fetal health during pregnancy, as it enables informed decisions about dietary choices impacting both mother and fetus. [16] further analyzes that despite evidence-based recommendations, limited health literacy poses challenges, leading to inadequate nutrient intake and delayed prenatal care, potentially risking neural tube defects and compromised maternal-fetal health.

In response to the challenges faced by pregnant women in receiving accurate and easy-to-understand nutrition advice, various tools and interventions have been developed to provide support and information. [14] discuss the role of platforms like Nutripedia in providing evidence-based nutritional information, aiming to combat misinformation and promote positive health behaviors. Similarly, [1] introduce chatbots of *Dr. Joy*, which offer continuous digital support to perinatal women and their partners. [14] suggest that ongoing research is necessary to address limitations and enhance the effectiveness of these interventions, particularly in leveraging advanced technologies such as LLMs in conversation.

2.2 Retrieval-Augmented Large Language Models

Several studies have delved into the exploration of RAG techniques and its implications across different domains. The study [4] explores the incorporation of

background explanation to enhance the accessibility of complex medical texts to a broader audience. By employing RAG techniques, the study investigates how external knowledge sources, such as definitions and explanations from resources like Wikipedia, can be utilized to augment the understanding of expert-authored content. The findings reveal the effectiveness of RAG in improving summary quality and simplicity while maintaining factual correctness. One of the challenges results indicated was that while these LLMs can generate simplified content, the quality of the summaries falls short of ideal standards.

Similarly, [11] explores the efficacy of RAG in incorporating external knowledge to enhance response generation within knowledge-intensive Natural Language Processing (NLP) tasks. It highlights the growing importance of integrating external knowledge sources into NLP systems for generating contextually relevant and informative responses. The effectiveness of RAG is evaluated through a comprehensive assessment that compares its design with current techniques and evaluates the retriever’s ability to select relevant passages. Through these evaluations, the study demonstrates RAG’s capability to utilize external information effectively, leading to state-of-the-art performance in knowledge-intensive NLP tasks. Another similar study [13] explores the application of RAG strategies to enhance the accuracy and relevance of LLMs in healthcare, particularly in nephrology. The research delves into the challenges posed by the imperfect accuracy and tendency to produce inaccuracies or irrelevant outputs inherent in LLMs, which are critical in medical applications requiring precision for decision-making processes. The study suggest the potential of RAG in enhancing the accuracy of specialized ChatGPT models tailored to align with specific medical guidelines. One such example includes the KDIGO 2023 guidelines, which provide guidance for chronic kidney disease.

The Work in Progress research [20] explored the potential of LLMs as a tool for providing personalized nutrition advice to pregnant patients with models such as ChatGPT demonstrating promising results. An illustrative example of this research is creation of NutritionBot, eliciting input and offering personalized nutrition recommendations based on demographic and personal history. Google Dialogflow is utilized to implement and host the chatbot, which integrates domain-specific nutrition questions and employs ChatGPT to generate contextually appropriate responses. Building upon the insights gained from [20], this research aims to further explore the efficacy of RAG in enhancing chatbot response quality, particularly in the context of nutrition advising for pregnant women, while also addressing limitations identified by previous research employing LLMs for response generation.

3 Method

3.1 Retrieval-Augmented Large Language Models

In order to develop a robust RAG system capable of providing personalized nutrition advice in a production-ready environment, meticulous design and implementation are essential. The proposed architecture is illustrated in Figure 1.

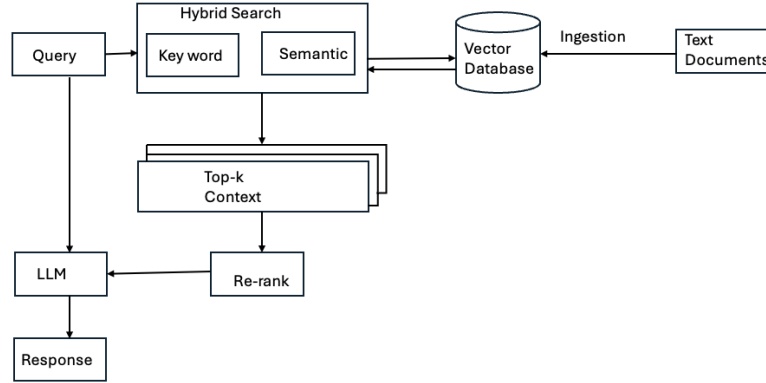


Fig. 1. RAG architecture

The initial crucial step involves the indexing process, wherein a vast amount of relevant nutritional data is systematically organized, processed, and stored in a structured format to facilitate efficient querying. This indexing phase lays the foundation for the subsequent retrieval or search process, enabling the RAG system to identify and retrieve pertinent information from the indexed database rapidly.

The retrieval or search process generates accurate and personalized nutrition advice within the RAG system. Upon receiving a user’s query or input, the RAG system swiftly searches the indexed database to retrieve relevant information, including nutritional facts, dietary recommendations, and health guidelines. The retrieved data is subsequently utilized by the generation component of the RAG system to produce coherent, informative, and tailored nutrition advice that addresses the user’s specific needs and concerns.

The indexing process commences with the segmentation of text into smaller, manageable chunks, which may be as granular as a sentence, paragraph, or specific sequence of 500 characters. Each chunk is then transformed into a vector of predetermined length using an embedding algorithm. The resultant vectors are subsequently stored in a vector database optimized for efficient retrieval.

The development of a RAG system critically depends on the ability to locate relevant contextual information. Retrieving pertinent documents or knowledge significantly aids the generation process. While traditional vector-based search methods can capture semantic relevance, they sometimes struggle with precise term matches and brief queries. To address these challenges, we investigate the implementation of hybrid search and rerank methods to improve retrieval effectiveness.

Hybrid search: In our RAG framework, hybrid search involves the concurrent use of vector and keyword searches. By establishing both vector and keyword indices in the database beforehand, the system capitalizes on the strengths of each method when processing a user query. This dual approach ensures a more

comprehensive and precise retrieval capability, thereby enhancing the quality of information supplied to the generative model.

Rerank method: The rerank method plays a pivotal role in refining the search process, thus increasing the relevance and accuracy of the generated responses. Empirical evidence suggests that employing a reranker can significantly improve the results of initial retrieval efforts. In this method, a hybrid search is conducted to retrieve a broad set of candidate documents. Each document is then assessed for its relevance to the query using a scoring mechanism that evaluates semantic alignment. Subsequently, documents are reordered based on their relevance scores, with the top K documents selected for final output. This procedure prioritizes the retrieval of semantically pertinent documents. For this process, we employ Cohere’s multilingual rerank model.

This study leverages the Dify framework, which incorporates Nomic’s advanced indexing techniques to embed textual chunks [17]. This framework offers a multifaceted approach, enabling not only the visualization of embedded vectors but also providing flexibility to experiment with diverse indexing configurations. Users can modify indexing methods, explore metadata details, and adjust retrieval counts, among other settings, thereby rendering it an indispensable tool in optimizing the indexing and retrieval processes.

The development of production-ready applications utilizing LLMs poses several challenges, including computational difficulties, data privacy concerns, integration complexities, latency issues, and others. However, the Dify framework facilitates the orchestration of LLMs in production environments by supporting various RAG features, ETL capabilities, Vector Databases, and other functionalities. In our experiments, we employed different sizes of the Llama 2 model, specifically 7B, 13B, and 70B, all quantized to a 4-bit length. We utilized Dify’s model for indexing and Cohere’s multilingual model for reranking. Our experimental setup consisted of a server equipped with two NVIDIA RTX4090 GPUs, 128 MB DDR5 RAM, and an Intel 14th Gen Core i9 CPU.

3.2 Dataset and Evaluations

The datasets used to construct the RAG model in this study were rigorously curated from diverse sources, including federal agencies and scientific publications, to establish a comprehensive knowledge base for nutrition-related factors during pregnancy. An initial review of 20 information sources commonly used to inform patients during office visits was conducted independently by three co-authors with expertise in healthcare and nutrition. This search was subsequently expanded to incorporate 20 online nutrition e-books, augmenting our dataset’s size. The primary datasets were further supplemented by collecting data from reputable sources, including the Food and Drug Administration (FDA), the Centers for Disease Control and Prevention (CDC), and academic papers indexed in the PubMed database. Ultimately, 12,000 words (comprising 40 documents) were indexed in the vector database.

To assess the effectiveness of the proposed model, we generated a dataset comprising 117 unique sample questions from ChatGPT-4. The prompt used

was “Provide 100 common questions that pregnant women frequently ask about nutrition.” This prompt was repeated twice, resulting in 117 distinct questions that were utilized as the evaluation dataset. These questions were subsequently submitted to various candidate models to elicit responses. The sample questions included “How much water should I drink while pregnant?”, “Are there specific nutrients that support the immune system during pregnancy?”, and “What nutritional considerations are there for women experiencing multiple pregnancies (twins, triplets, etc.)?” among others.

The evaluation of the responses from the candidate models focused on two key aspects: general quality and specificity to the posed questions. To facilitate this assessment, the following prompt template was used to rate the quality of the responses. This approach aimed to measure both the accuracy and the applicability of the responses provided by the models. We modified the definition of quality to assess scores from these two aspects. The definition of general quality is presented below.

Prompt: On a scale from 1 to 100, where 1 represents the lowest quality and 100 signifies the highest quality, please evaluate the quality of the response provided to a user. The user’s description is enclosed within `< user >< /user >` tags, the question posed by the user within `< question >< /question >` tags, and the corresponding answer within `< answer >< /answer >` tags. *Quality is defined by the answer’s reliability, conciseness, and clarity, ensuring it conveys the necessary information without superfluous details or misleading the user.* Please provide only the scale, no explanation or other text.

```
< user >The user is a pregnant woman seeking nutrition advice. <
/user >
<question>How much water should I drink while pregnant?< /question >
< answer > During pregnancy, it’s important to stay properly hy-
drated to support your health and your baby’s growth. The recom-
mended daily intake of water for pregnant women is eight to twelve
8-ounce glasses, which can increase depending on factors such as phys-
ical activity, hot weather, or fever. Proper hydration can help reduce
swelling, constipation, and fatigue, and improve digestion, elimination,
and overall well-being. It’s best to drink water regularly throughout the
day, aiming for at least 50 ounces of fluid intake from water, milk, juice,
and other beverages. You can also monitor your urine output to ensure
you’re drinking enough - if it’s light yellow or clear, you’re on the right
track!< /answer >
```

We revised the definition of specificity to state: “*Quality is defined by the answer’s specificity. The shorter and easier-to-read answers should receive higher scores. Answers should be summarized from the context, rather than presented as bullet points.*”

In this study, we submitted a comprehensive set of prompts to the OpenAI API and retrieved the corresponding quality scores. Our experimental design

involved evaluating the performance of different LLaMA2 model sizes (7B, 13B, and 70B) with and without embedding context, resulting in a total of 702 prompts being sent for evaluation. To ensure the accuracy and reliability of our results, we employed both ChatGPT 3.5 and 4 models in our assessment ³. A paired t-test was conducted to compare the mean values of the returned quality scores between models of the same size.

4 Findings

The results of our experiment are presented in Tables 1 and 2. Our analysis indicates that the RAG models consistently produced significantly shorter responses than the baseline models, a trend observed across all three model sizes. Furthermore, incorporating context into the prompt via the extrinsic result tended to yield more concise responses. Specifically, the context embedding approach generated responses that were approximately 50% shorter in length. Below is a sample response to the prompt 'How much water should I drink while pregnant?'

LLAMA-2 70b with an embedded RAG context During pregnancy, it's important to stay properly hydrated to support your health and your baby's growth. The recommended daily intake of water for pregnant women is eight to twelve 8-ounce glasses, which can increase depending on factors such as physical activity, hot weather, or fever. Proper hydration can help reduce swelling, constipation, and fatigue, and improve digestion, elimination, and overall well-being. It's best to drink water regularly throughout the day, aiming for at least 50 ounces of fluid intake from water, milk, juice, and other beverages. You can also monitor your urine output to ensure you're drinking enough - if it's light yellow or clear, you're on the right track!

LLAMA-2 70b without an embedded RAG context It's important to stay well-hydrated during pregnancy, as your body will be working hard to support the growth of your baby. The amount of water you should drink while pregnant will depend on several factors, including your age, weight, activity level, and the climate you live in.

The Institute of Medicine (IOM) recommends that pregnant women drink at least 8-10 cups (64-80 ounces) of water per day. However, some healthcare providers may recommend more or less water depending on individual needs.

In addition to plain water, you can also consume water-rich foods such as fruits, vegetables, and broth-based soups to help meet your fluid needs. It's also important to note that you can overhydrate, so it's important to listen to your body and drink when you feel thirsty rather than forcing yourself to drink a certain amount of water each day.

³ All experiments were completed in April 2024. We used the latest version of the ChatGPT subscription and the APIs.

It’s also important to note that some women may experience swelling or edema during pregnancy, which can be a sign of excess fluid retention in the body. In this case, your healthcare provider may recommend limiting your fluid intake to help manage the swelling.

Overall, it’s important to work with your healthcare provider to determine the right amount of water for you during pregnancy. They can help you assess your individual needs and make recommendations based on your specific situation.

Table 1. Length of Response (Number of Characters)

Model	Context M(SD)	Baseline M(SD)	
Llama2-7b	1227.27 (518.39)	2548.75 (496.83)	**
Llama2-13b	1373.79 (582.46)	2367 (476.10)	**
Llama2-70b	1202.55 (571.24)	2289.92 (615.57)	**

*Mean (M), Standard Deviation (SD), * $p < 0.05$, ** $p < 0.01$*

Table 2. The results of general quality & specificity test

	Model	ChatGPT 3.5			ChatGPT 4		
		RAG M(SD)	Baseline M(SD)		RAG M(SD)	Baseline M(SD)	
Quality	7b	78.93(12.55)	82.26(11.01)	*	95.12(8.81)	97.90(2.87)	**
	13b	76.28(13.94)	83.27(9.80)	**	95.25(9.80)	98.29(4.41)	**
	70b	79.35(12.79)	83.37(8.85)	**	96.02(5.23)	98.84(2.21)	**
Specificity	7b	52.86(22.28)	37.54(16.06)	**	81.66(14.18)	64.10(14.39)	**
	13b	51.06(21.44)	41.36(18.14)	**	80.72(13.93)	67.26(13.99)	**
	70b	56.40(23.29)	41.15(19.35)	**	85.51(9.59)	66.79(15.95)	**

In our exhaustive assessment of model performance, we observed a consistent pattern of superior performance by the 70b model compared to smaller models, including the 7b and 13b models. Notably, the 13b model exhibited enhanced performance relative to the 7b model when evaluated using ChatGPT 3.5, although this advantage was not replicated with ChatGPT 4. Across all candidate models, our results demonstrate that baseline models consistently outperformed RAG models in terms of overall quality, a finding supported by statistical testing confirming these differences’ significance. Furthermore, our analysis suggests that longer and more informative responses tend to receive higher rating scores for general quality, likely due to their increased content density, as exemplified by the sample response provided above.

In our evaluation of specificity quality, we observed that the 70b model with RAG embedding outperformed the 7b and 13b models. Conversely, the 13b model emerged as the top performer in the baseline model, surpassing the larger 70b model. This finding suggests that smaller models can also generate high-quality responses that rival those of larger models when equipped with contextualized embeddings. Moreover, our results reveal a significant effect of RAG embedding, with the enhanced model outperforming the baseline model in generating shorter responses to questions. These findings imply that incorporating contextual information through embeddings can lead to more concise, accurate, and informative responses.

5 Conclusions

This study investigates whether Large Language Models (LLMs) augmented with Retrieval-Augmented Generation (RAG) can enhance the quality of responses provided in pregnancy nutrition advice. Our experimental results reveal that RAG models consistently generated shorter responses compared to baseline models across all model sizes, with context embedding yielding even more concise responses (approximately 50% shorter). Notably, the 70b model outperformed smaller models in terms of overall response quality, whereas the 13b model demonstrated superior performance when evaluated using ChatGPT 3.5. Furthermore, our analysis shows that the 70b model with RAG embedding excelled in specificity quality, while the 13b baseline model surprisingly surpassed the larger 70b model. These findings suggest that incorporating contextual information through embeddings could lead to more concise, accurate, and informative responses and that smaller models can still produce high-quality responses when equipped with contextualized embeddings.

Reading and comprehending health information is a fundamental skill for patients to make informed decisions about their care [18]. In healthcare settings, healthcare providers, educators, and patients often prefer receiving concise and easily digestible information. Our experimental results suggest that RAG-enhanced LLMs can generate concise information that may be more accessible to users. This finding sheds light on the potential for adopting this technology in healthcare to improve patient health literacy. In our context, a pregnant woman could utilize this tool to access and understand information about nutrition during pregnancy. Based on doctor-approved data as context, the model could provide useful responses to her queries at any time and from any location. This highlights a novel interaction and use case for pregnant women to improve nutritional advice during pregnancy.

Our experimental results suggest that, in constrained contexts, smaller models can provide similar or even superior responses to users when the context is properly embedded. This finding has significant implications, as larger models require substantial computing power (i.e., powerful GPUs), which can be a major accessibility issue for users in rural areas or those with limited access to high-performance machines or the financial means to afford such equipment. In

contrast, smaller models like 7B could be executed on regular personal computers or mobile devices, thereby increasing accessibility to more users and leveraging the power of large language models. Further studies could continue to explore the area of edge computing and its deployment to end-users.

We plan to address several limitations of this study in future research endeavors. One key aspect crucial to ensuring the success of RAG embedding is data quality, which was compromised in this experiment due to the limited dataset of only 40 documents collected as proof-of-concept. A more comprehensive dataset would be necessary to test the effectiveness of RAG embedding further. Another essential consideration is indexing performance, a critical factor in RAG that requires proper contextualization to avoid reproducing results from the original model. Unfortunately, we did not monitor the performance of indexing in this study, which constitutes a limitation. Furthermore, our evaluation relied on a GPT-powered assessment to test information quality, which may not guarantee quality for healthcare providers and patients. These issues necessitate further exploration and research to realize the full potential of RAG embedding.

References

- [1] Kyungmi Chung, Hee Young Cho, and Jin Young Park. “A chatbot for perinatal women’s and partners’ obstetric and mental health care: Development and usability evaluation study”. In: *JMIR Medical Informatics* 9.3 (2021), e18607.
- [2] Luciano Floridi and Massimo Chiriatti. “GPT-3: Its nature, scope, limits, and consequences”. In: *Minds and Machines* 30 (2020), pp. 681–694.
- [3] Nayara Gomes Graciliano, Jonas Augusto Cardoso da Silveira, and Alane Cabral Menezes de Oliveira. “Consumo de alimentos ultraprocessados reduz a qualidade global da dieta de gestantes”. In: *Cadernos de Saúde Pública* 37.2 (2021), e00030120.
- [4] Yue Guo et al. “Retrieval augmentation of large language models for lay language generation”. In: *Journal of Biomedical Informatics* 149 (2024), p. 104580.
- [5] Kai He et al. “A survey of large language models for healthcare: from data, technology, and applications to accountability and ethics”. In: *arXiv preprint arXiv:2310.05694* (2023).
- [6] Zeyu He et al. “If in a Crowdsourced Data Annotation Pipeline, a GPT-4”. In: *arXiv preprint arXiv:2402.16795* (2024).
- [7] Fei Jiang et al. “Artificial intelligence in healthcare: past, present and future”. In: *Stroke and vascular neurology* 2.4 (2017).
- [8] Sarah Louise Killeen et al. “Using FIGO Nutrition Checklist counselling in pregnancy: A review to support healthcare professionals”. In: *International Journal of Gynecology & Obstetrics* 160 (2023), pp. 10–21.
- [9] Berthold Koletzko et al. “Nutrition during pregnancy, lactation and early childhood and its implications for maternal and long-term child health:

- the early nutrition project recommendations”. In: *Annals of Nutrition and Metabolism* 74.2 (2019), pp. 93–106.
- [10] Simone Kresevic et al. “Optimization of hepatological clinical guidelines interpretation by large language models: a retrieval augmented generation-based framework”. In: *npj Digital Medicine* 7.1 (2024), p. 102.
 - [11] Patrick Lewis et al. “Retrieval-augmented generation for knowledge-intensive nlp tasks”. In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 9459–9474.
 - [12] Nicole E Marshall et al. “The importance of nutrition in pregnancy and lactation: lifelong consequences”. In: *American journal of obstetrics and gynecology* 226.5 (2022), pp. 607–632.
 - [13] Jing Miao et al. “Integrating retrieval-augmented generation with large language models in nephrology: advancing practical applications”. In: *Medicina* 60.3 (2024), p. 445.
 - [14] João Luis Zeni Montenegro, Cristiano André da Costa, and Luisa Plácido Janssen. “Evaluating the use of chatbot during pregnancy: A usability study”. In: *Healthcare Analytics* 2 (2022), p. 100072.
 - [15] Afsana Naaz and Komal N Muneshwar. “How Maternal Nutritional and Mental Health Affects Child Health During Pregnancy: A Narrative Review”. In: *Cureus* 15.11 (2023).
 - [16] Farah Nawabi et al. “Health literacy in pregnant women: a systematic review”. In: *International journal of environmental research and public health* 18.7 (2021), p. 3847.
 - [17] Zach Nussbaum et al. *Nomic Embed: Training a Reproducible Long Context Text Embedder*. 2024. arXiv: 2402.01613 [cs.CL].
 - [18] Rima E Rudd, Lindsay Rosenfeld, and Vanessa W Simonds. “Health literacy: a new area of research with links to communication”. In: *Atlantic Journal of Communication* 20.1 (2012), pp. 16–30.
 - [19] Jamie V de Seymour, Kathryn L Beck, and Cathryn A Conlon. “Nutrition in pregnancy”. In: *Obstetrics, Gynaecology & Reproductive Medicine* 29.8 (2019), pp. 219–224.
 - [20] Chun-Hua Tsai et al. “Generating Personalized Pregnancy Nutrition Recommendations with GPT-Powered AI Chatbot. In: 20th International Conference on Information Systems for Crisis Response and Management”. In: *20th International Conference on Information Systems for Crisis Response and Management (ISCRAM)*. Vol. 2023. 2023, p. 263.
 - [21] Cara A Yelverton et al. “Diet and mental health in pregnancy: Nutrients of importance based on large observational cohort data”. In: *Nutrition* 96 (2022), p. 111582.