ChatGPT: An Evaluation of AI-Generated Responses to Commonly Asked Pregnancy Questions

Christopher Wan, Angelo Cadiente, Keren Khromchenko, Natalie Friedricks, Rima A. Rana, Jonathan D. Baum*

Department of Obstetrics and Gynecology, Jersey Shore University Medical Center, Neptune, NJ, USA
Email: *jonathan.baum@hmhn.org

Abstract

Background: A recent assessment of ChatGPT on a variety of obstetric and gynecologic topics was very encouraging. However, its ability to respond to commonly asked pregnancy questions is unknown. Reference verification needs to be examined as well. **Purpose:** To evaluate ChatGPT as a source of information for commonly asked pregnancy questions and to verify the references it provides. **Methods:** Qualitative analysis of ChatGPT was performed. We queried ChatGPT Version 3.5 on 12 commonly asked pregnancy questions and asked for its references. Query responses were graded as "acceptable" or "not acceptable" based on correctness and completeness in comparison to American College of Obstetricians and Gynecologists (ACOG) publications, PubMed-indexed evidence, and clinical experience. References were classified as "verified", "broken", "irrelevant", "non-existent" or "no references". Review and grading of responses and references were performed by the co-authors individually and then as a group to formulate a consensus. **Results:** In our assessment, a grade of acceptable was given to 50% of responses (6 out of 12 questions). A grade of not acceptable was assigned to the remaining 50% of responses (5 were incomplete and 1 was incorrect). In regard to references, 58% (7 out of 12) had deficiencies (5 had no references, 1 had a broken reference, and 1 non-existent reference was provided). **Conclusion:** Our evaluation of ChatGPT confirms prior concerns regarding both content and references. While AI has enormous potential, it must be carefully evaluated before being accepted as accurate and reliable for this purpose.

Keywords

AI (Artificial Intelligence), ChatGPT, Pregnancy

1. Introduction

Artificial intelligence (AI) has emerged as a transformative force and expanded into everyday life. Large-language models (LLM), a subset of AI, have rapidly become popular due to their near-human ability to write coherently across an endless range of topics. Chat Generative Pre-Training Transformer (ChatGPT) Version 3.5 (also referred to as GPT-3.5 by its creator, OpenAI) is the fastest growing internet application to reach 100 million users, achieving this feat within 2 months of its release in November 2022 [1]. ChatGPT represents a technological advancement with the ability to generate a refined, organic response to specific questions in a manner that a traditional search engine cannot.

ChatGPT has shown tremendous promise in clinical application. It demonstrated competence in clinical reasoning and general medical knowledge by passing USMLE Steps 1, 2, and 3 without formal medical training [2] [3]. It outperformed physicians in regard to quality and empathy questions, suggesting a potential future role in patient messaging [4]. Grünebaum *et al.* (2023) assessed ChatGPT responses with a variety of questions about obstetrics and gynecology (queried in February 2023) and concluded that this LLM has potential to provide information about "virtually any topic in obstetrics and gynecology". They described ChatGPT's responses to be "nuanced, eloquent, [and] informed". They acknowledged, however, it has the "potential to mislead" due to an "apparent lack of insight", categorizing the chatbot as "a work in progress" with the potential to cause patient harm [5].

ChatGPT's matter-of-fact responses are not without flaw. It may hallucinate output or "make up facts". AI-generated essays submitted by college students have been described as "really well-written [but] wrong" [6]. In other words, ChatGPT has the capacity to convincingly present misinformation. In regard to references, ChatGPT may incorrectly list authors, articles, and/or PubMed identifiers when asked for its sources [7]. Additionally, the foundation of indexed information used to train the most accessible version, GPT-3.5, is limited to data preceding November 2021 [8]. Despite these limitations, ChatGPT remains an enticing tool that can process and present information in a way that would otherwise not be possible with an internet search engine alone. This chatbot is an aspiring middle-man between the world's knowledge and the user. Its impact is too inevitable to ignore.

We know that pregnant women frequently use the internet for information regarding pregnancy topics. A survey of postpartum women with internet access estimated that more than 80% use it at least weekly to answer their pregnancy-related questions [9]. The cat is out of the proverbial bag: given their already massive user bases and inexorable expansion, it is the responsibility of subject experts to verify the accuracy of the information that LLMs provide. This is the logical initial step in the evaluation of any new tool.

We acknowledge that a *caveat emptor* approach makes sense to a panel of reviewers with postgraduate degrees; however, use by individuals without this ad-

vanced training may be problematic. Popp (2023) suggests that "users need to double check absolutely everything" [10]. While it is reasonable to assume that subject experts will view ChatGPT output with sufficient skepticism, it is unlikely that the average end-user will do the same.

AI technology is expected to infiltrate most industries within the next decade prompting many to voice concerns regarding human replacement by these 21st century machines. While the interest and study of chatbots continues to surge rapidly, the issue of user confidence and acceptance of this technology remains unclear [10] [11].

The purpose of this study is to evaluate ChatGPT as a source of acceptable answers for commonly asked pregnancy questions and to verify the references it provides. Proof that ChatGPT is a valid and trustworthy source of pregnancy information based on reliable references is essential prior to its endorsement for this purpose.

For a glossary of terminology relating to artificial intelligence, please see the Appendix section.

2. Methods

We queried ChatGPT-3.5 with 12 commonly asked pregnancy questions based on current literature and clinical experience. We chose the basic format "Can I [action] in pregnancy?" as we anticipate a pregnant woman might do when seeking information regarding a specific activity as it relates to her pregnancy. We excluded questions that were deemed too broad, such as "What medications are safe in pregnancy?" "Can I exercise during pregnancy?" and "Can I go to the dentist while pregnant?" Each queried question was followed up by the prompt "Tell me your references for this question" to record the LLM's source for its response. A new chat was created each time with the prior data unsaved to avoid the model's learning mechanism. All queries were conducted and recorded from May 14-20, 2023. The query questions are listed in Table 1.

Table 1. Common pregnancy questions used to query ChatGPT.

1) Can I fly on a plane while pregnant?

2) Can I drink alcohol while pregnant?

3) Can I drink coffee while pregnant?

4) Can I eat deli meats while pregnant?

5) Can I dye my hair while pregnant?

6) Can I do laser hair removal while pregnant?

7) Is marijuana use safe during pregnancy?

8) Can I eat sushi while pregnant?

9) Can I vape while pregnant?

10) Can I do yoga while pregnant?

11) Can I go in a hot tub while pregnant?

12) What sleep position is best during pregnancy?

Responses were recorded, reviewed and graded by each co-author individually and then as a group to formulate a consensus. The review team consisted of two board-certified OBGYN physicians (JDB, RAR), two OBGYN resident physicians (KK, NF), and two second year medical students (CW, AC) of which three identified as female, three as male with an age range from 24 to 56 years old. We chose American College of Obstetricians and Gynecologists (ACOG) Practice Bulletins and Committee Opinions as the default expert opinion (whenever available) against which ChatGPT responses were compared. When ACOG guidance was not available, we used best available PubMed-indexed evidence in addition to clinical experience. Each response was evaluated and assigned a grade as defined in Table 2. Each reference provided was evaluated using the criteria listed in Table 3.

3. Results

ChatGPT was queried on 12 commonly asked pregnancy questions and was asked for references. Its verbatim response in italics, consensus grade, and grade explanation are listed in that order. Summaries of ChatGPT Response Grading and ChatGPT Reference Evaluation are presented in Table 4 and Table 5, respectively.

Table 2. ChatGPT response grading scale (adapted from Yeo 2023 with permission) [12].

Grade	Interpretation
Acceptable	Response is free from error and contains all necessary information.
Not Acceptable	Response is missing necessary information (**Incomplete**).
	Response contains factual errors (**Incorrect**).

Table 3. ChatGPT reference criteria.

Reference Issue	Explanation
Broken Reference	ChatGPT provided a website address or an article DOI number but the website's page led to an error message and/or the DOI number was misleading. Digital Object Identifier, DOI, is a string of numbers, letters, or symbols used to uniquely identify an article or document on the internet.
Irrelevant Reference	ChatGPT provided a list of references for a query that had little or nothing to do with the topic of interest.
Non-Existent Reference	ChatGPT listed references that do not exist.
No References	ChatGPT could not provide a list of references.
Verified Reference	ChatGPT provided verified reference(s).

Table 4. ChatGPT response grading.

Question	Response Grade	Explanation
Air travel	Not Acceptable (Incomplete)	ACOG states, "Most commercial airlines allow pregnant women to fly up to 36 weeks of gestation [13]." ChatGPT failed to mention this restriction.
Alcohol	Acceptable	Both ACOG and ChatGPT strongly recommend against the use of alcohol [19].
Caffeine	Acceptable	Consistent with ACOG and provides a dose-limited recommendation along with various sources of caffeine [20].
Deli Meat	Acceptable	Consistent with ACOG recommendations on Listeria and proper food safety guidelines [21].
Hair Dye	Acceptable	Consistent with ACOG expert opinion on dying hair [22].
Laser Hair Removal	Not Acceptable (Incomplete)	ChatGPT fails to consider the following evidence from a systematic review from 2019. "The available evidence, limited to low evidence-level case reports and series, indicates cutaneous laser treatment during pregnancy is safe for both mother and fetus [14]."
Marijuana Use	Acceptable	ACOG and ChatGPT both recommend against marijuana use, while acknowledging potential risks and uncertainty of current evidence [23].
Sushi	Not Acceptable (Incorrect)	ChatGPT incorrectly states that women can eat sushi during pregnancy. According to ACOG, "although not mentioned in the revised 2017 advice, it is important that pregnant women avoid all raw and undercooked seafood, eggs, and meat [15]."
Vaping	Not Acceptable (Incomplete)	ChatGPT fails to mention the CDC's advisory statement regarding vaping in pregnancy. "Recently, the Centers for Disease Control and Prevention (CDC) issued an advisory notice investigating a multistate outbreak of noninfectious severe pulmonary disease associated with e-cigarette and vaping product use. With the recent CDC advisory and the effects of e-cigarette and vaping product use on offspring health, immediate discontinuation of e-cigarette and vaping products should be advised among all pregnant and postpartum women [16]."
Yoga	Not Acceptable (Incomplete)	ChatGPT does not include prospective data from a study published in 2015. Twenty-six yoga positions in healthy pregnant women in their third trimester were evaluated. Maternal and fetal status were evaluated during the yoga session. There were no adverse changes in maternal and fetal well-being during the session [17].
Hot Tub	Acceptable	Consistent with ACOG and AAP, ChatGPT cautions regarding hot tub exposure and ways to reduce risk of harmful fetal effects [24].
Sleep Position	Not Acceptable (Incomplete)	ChatGPT does not include information from a prospective study on maternal sleeping position published in 2019 which concluded "supine or non-left sided sleep through 30 weeks gestation was not associated with adverse pregnancy outcomes linked to decreased placental blood flow in a large prospective cohort [18]."

Table 5. ChatGPT reference evaluation.

Question	Reference Issue
Air Travel	Broken Reference Irrelevant Reference
Alcohol	No Reference
Caffeine	Verified Reference
Deli Meat	Verified Reference
Hair Dye	No Reference
Laser	Verified Reference
Marijuana	Verified Reference
Sushi	Non-Existent Reference
Vaping	No Reference
Yoga	No Reference
Hot Tub	No Reference
Sleep Position	Verified Reference

3.1. Questions, ChatGPT Response, Grade, and Explanation

3.1.1. Can I Fly on a Plane While Pregnant?

In general, flying on a plane while pregnant is considered safe for most women, but it is recommended that you speak with your healthcare provider before making any travel plans. Your doctor can assess your individual pregnancy and any potential risks or complications that may arise from air travel.

Some airlines have specific policies related to pregnant passengers, so it's important to check with your airline to see if they have any restrictions or requirements. Additionally, it's a good idea to plan for extra comfort during your flight by booking an aisle seat, bringing plenty of water, and wearing loose, comfortable clothing.

It's also important to keep in mind that the risk of developing blood clots during long flights may be increased for pregnant women. To reduce this risk, you can try to move around regularly during the flight, drink plenty of water, and wear compression stockings.

In summary, it is generally safe for most pregnant women to fly on a plane, but it is important to check with your doctor and the airline before making any travel plans.

GRADE: Not Acceptable (Incomplete)

We found ChatGPT's air travel response to be not acceptable because it is incomplete. Its omission of the statement that "most commercial airlines allow pregnant women to fly up to 36 weeks of gestation" could influence the purchase of non-refundable tickets or plainly lead to unsafe air travel after 36 weeks. In addition, the response lacks recommendations against air travel for women with medical or obstetric conditions that could be exacerbated by flight or that could

require emergency care [13].

3.1.2. Can I Drink Alcohol While Pregnant?

As an AI language model, I must prioritize your health and the well-being of your baby. It is strongly recommended to avoid drinking alcohol during pregnancy. Consuming alcohol during pregnancy can pose significant risks to the developing fetus, leading to a condition known as fetal alcohol spectrum disorders (FASDs).

FASDs encompass a range of physical, behavioral, and cognitive disabilities that can occur in babies exposed to alcohol in the womb. These effects can be permanent and have lifelong consequences. There is no known safe level of alcohol consumption during pregnancy, so it's best to err on the side of caution and abstain from drinking alcohol entirely while pregnant.

If you have concerns or questions about alcohol consumption during pregnancy, it's crucial to consult with your healthcare provider, who can provide personalized advice and guidance based on your specific circumstances.

GRADE: Acceptable

3.1.3. Can I Drink Coffee While Pregnant?

The consumption of coffee during pregnancy is a topic of debate among healthcare professionals. While moderate caffeine intake is generally considered safe during pregnancy, excessive consumption of caffeine has been associated with an increased risk of certain adverse pregnancy outcomes, such as miscarriage, preterm birth, and low birth weight.

The American College of Obstetricians and Gynecologists recommends that pregnant women limit their caffeine intake to no more than 200 milligrams per day, which is roughly equivalent to one 12-ounce cup of coffee. It's important to note that caffeine is not just found in coffee, but also in other beverages such as tea, soda, and energy drinks, as well as in some foods and medications.

It's always a good idea to discuss your caffeine intake with your healthcare provider to determine what level of consumption is safe and appropriate for you during pregnancy. If you choose to drink coffee, you may want to consider switching to decaf or limiting your consumption to one small cup per day to help minimize any potential risks.

GRADE: Acceptable

3.1.4. Can I Eat Deli Meats While Pregnant?

While it is generally safe to eat deli meats during pregnancy, there are some precautions that you should take to reduce the risk of getting an infection called listeriosis. Listeriosis is caused by the bacterium Listeria monocytogenes, which can be found in some types of deli meats.

To reduce the risk of listeriosis, you should follow these guidelines:

1) Heat deli meats until they are steaming hot before eating them. This will kill any Listeria bacteria that may be present.

2) Avoid eating cold deli meats or meat spreads from the deli counter, unless

they have been heated to steaming hot just before serving.

3) Avoid eating refrigerated pâté or meat spreads, as they can contain Listeria.

4) Always wash your hands before and after handling deli meats, and clean any surfaces that come into contact with deli meats.

By taking these precautions, you can safely enjoy deli meats during pregnancy. However, if you are unsure about the safety of any food, it is always best to consult with your healthcare provider.

GRADE: Acceptable

3.1.5. Can I Dye My Hair While Pregnant?

Dyeing your hair while pregnant is a topic that has been a subject of debate among medical professionals. The concern revolves around the potential risks associated with the chemicals present in hair dyes. While limited research is available on the topic, it is generally considered safe to dye your hair during pregnancy, especially after the first trimester. However, it's always recommended to consult with your healthcare provider to get personalized advice based on your specific situation.

If you decide to dye your hair while pregnant, consider the following precautions:

1) Opt for safer alternatives: Look for hair dyes that are ammonia-free and contain fewer harsh chemicals. Additionally, semi-permanent or vegetable-based dyes are often considered safer options.

2) Good ventilation: Make sure you're in a well-ventilated area when applying the hair dye to minimize exposure to any fumes.

3) Wear gloves: Use gloves to protect your skin and minimize contact with the dye.

4) Timing: It's generally advisable to wait until the second trimester to dye your hair. This is because the first trimester is a crucial period of development for the baby.

5) Patch test: Perform a patch test on a small area of your skin before applying the dye to your entire scalp. This can help identify any potential allergic reactions.

6) Limit scalp exposure: Try to minimize the amount of dye that comes into contact with your scalp, as it may be more easily absorbed into your system.

Remember, every pregnancy is unique, and individual circumstances may vary. Therefore, it's essential to consult with your healthcare provider before making a decision about dyeing your hair while pregnant. They will be able to provide you with personalized advice based on your medical history and current condition.

GRADE: Acceptable

3.1.6. Can I Do Laser Hair Removal While Pregnant?

It is generally not recommended to undergo laser hair removal while pregnant. Although there is no concrete evidence that laser hair removal is harmful to the

fetus, there have not been enough studies done to determine the safety of the procedure during pregnancy.

Additionally, pregnancy can cause changes in your skin, including increased sensitivity and pigmentation changes, which can make laser hair removal less effective or potentially cause adverse effects.

It is important to prioritize the health and safety of both you and your baby during pregnancy, so it is best to avoid unnecessary risks. If you are considering laser hair removal, it is best to wait until after you have given birth and are no longer breastfeeding. In the meantime, there are other hair removal methods, such as shaving or using depilatory creams, that are safe during pregnancy.

GRADE: Not Acceptable (Incomplete)

We found ChatGPT's laser hair removal response to be not acceptable because it is incomplete. Its omission of a systematic review from 2019 on the safety of cutaneous laser treatment during pregnancy has the potential to deprive women of the necessary information upon which an informed decision regarding laser hair removal may be made [14].

3.1.7. Is Marijuana Use Safe during Pregnancy?

The safety of marijuana use during pregnancy is a topic of ongoing debate and research. While it is difficult to draw definitive conclusions due to limited and conflicting studies, the general consensus among healthcare professionals is that using marijuana during pregnancy is not recommended.

Marijuana contains compounds called cannabinoids, including delta-9-tetra-hydrocannabinol (THC), which can cross the placenta and affect the developing fetus. Some studies suggest that prenatal marijuana exposure may be associated with an increased risk of negative outcomes, such as low birth weight, preterm birth, and developmental and behavioral issues in children. However, it's important to note that research in this area is still evolving, and more conclusive evidence is needed.

Additionally, marijuana use during pregnancy may also be associated with other potential risks. Smoking marijuana, for example, exposes both the mother and the fetus to the harmful effects of smoke, which can be harmful to lung health. Furthermore, marijuana use can impair cognitive function and attention, which may impact a pregnant woman's ability to care for herself and her developing baby.

Given the uncertainty and potential risks, it is generally advised that pregnant women avoid using marijuana or any other recreational drugs during pregnancy. If you have concerns or questions about drug use during pregnancy, it is best to consult with a healthcare professional who can provide personalized guidance based on your specific situation.

GRADE: Acceptable

3.1.8. Can I Eat Sushi While Pregnant?

Yes, you can eat sushi while pregnant, but you need to be cautious and take cer-

tain precautions to ensure that the sushi is safe to consume. Raw fish used in sushi can potentially be contaminated with bacteria, viruses, or parasites, which can cause foodborne illnesses.

To reduce the risk of foodborne illness, you should make sure that the sushi is made with fresh, high-quality ingredients and prepared in a clean and sanitary environment. You should also avoid eating high-mercury fish, such as swordfish, shark, king mackerel, and tilefish, as these can be harmful to your developing baby.

It's important to talk to your healthcare provider about your diet during pregnancy and get their advice on the types and amounts of fish that are safe to eat. They can also advise you on any other dietary restrictions you may need to follow during pregnancy.

GRADE: Not Acceptable (Incorrect)

We found ChatGPT's sushi response to be not acceptable because it is incorrect. It states that women can eat sushi during pregnancy which *contradicts* ACOG's clear statement that pregnant women should avoid all raw and undercooked seafood, eggs and meat [15].

3.1.9. Can I Vape While Pregnant?

As an AI language model, I'm not a doctor, but I can provide you with some general information. It is generally recommended that pregnant women avoid using any form of nicotine or tobacco products, including vaping. Nicotine is a highly addictive substance that can have harmful effects on both the pregnant person and the developing fetus.

Vaping during pregnancy can expose the fetus to nicotine, as well as other potentially harmful chemicals found in e-cigarette liquids. Nicotine can constrict blood vessels and restrict the oxygen and nutrient supply to the fetus, potentially affecting its growth and development. It has been associated with various risks, including preterm birth, low birth weight, and developmental issues.

Additionally, vaping devices and e-cigarettes may malfunction, causing injury. The long-term effects of vaping on both the pregnant person and the fetus are still being researched, and the lack of conclusive evidence regarding the safety of vaping during pregnancy means it is generally best to err on the side of caution.

If you're pregnant and considering using any nicotine or tobacco products, it is advisable to consult with your healthcare provider for personalized advice. They can provide you with the most up-to-date and accurate information regarding the potential risks and alternatives to ensure the health and well-being of you and your baby.

GRADE: Not Acceptable (Incomplete)

We found ChatGPT's vape response to be not acceptable because it is incomplete. Its omission of the CDC's advisory statement regarding vaping in pregnancy has the potential to mislead. The CDC clearly recommends the discontinuation of e-cigarette and vaping products in all pregnant and postpartum women. This stance is supported by ACOG [16].

3.1.10. Can I Do Yoga While Pregnant?

Yes, yoga can be a safe and beneficial form of exercise during pregnancy, but it is important to take certain precautions and modifications to ensure the safety of both you and your growing baby.

It is recommended to consult with your healthcare provider before starting any new exercise program, including yoga. They can provide specific advice based on your individual pregnancy and any potential risks or complications.

When practicing yoga during pregnancy, it is important to avoid certain poses that may put undue stress on the abdomen, such as deep twists, inversions, or poses that involve lying on the stomach. It is also important to avoid overstretching, and to listen to your body and modify poses as needed.

There are many prenatal yoga classes and videos available that are specifically designed for pregnant women, with modified poses and sequences that are safe and effective for pregnancy. These classes may also include breathing techniques and relaxation practices that can help prepare you for labor and delivery.

Overall, with the proper modifications and guidance, yoga can be a beneficial form of exercise during pregnancy that can help improve flexibility, strength, and relaxation.

GRADE: Not Acceptable (Incomplete)

We found ChatGPT's yoga response to be not acceptable because it is incomplete. Its omission of prospective data from 2015 may lead a woman away from a potentially beneficial exercise activity shown to be safe during pregnancy [17].

3.1.11. Can I Go in a Hot Tub While Pregnant?

Before engaging in any activities during pregnancy, it's always best to consult with your healthcare provider for personalized advice. However, I can provide you with some general information about hot tub use during pregnancy.

Hot tubs, jacuzzis, or saunas that maintain a high temperature can potentially pose risks to pregnant women. Spending extended periods in hot water, especially if the water temperature exceeds 100˚F (38˚C), may lead to an increased body temperature, which can be harmful to the developing fetus.

Raising your core body temperature, particularly during the first trimester, has been associated with an increased risk of certain birth defects. It's also important to note that high water temperatures can potentially cause dehydration, which is not ideal during pregnancy.

If you still wish to use a hot tub during pregnancy, it's generally recommended to follow these precautions:

1) Limit your time: Keep your hot tub sessions short, typically less than 10 minutes at a time.

2) Lower the temperature: Make sure the water temperature is below 100˚F (38˚C). Ideally, it should be around body temperature or slightly below.

3) Stay hydrated: Drink plenty of water before and after using the hot tub to prevent dehydration.

4) Pay attention to warning signs: If you start feeling dizzy, lightheaded, or

overheated, it's important to exit the hot tub immediately and cool down.

5) *Avoid submerging your entire body*: To minimize the risk of overheating, you may choose to sit on the edge of the hot tub or keep your lower body in the water while allowing your upper body to remain exposed to cooler air.

Remember, every pregnancy is unique, and certain medical conditions or complications may make hot tub use more risky. Always consult with your healthcare provider to ensure that hot tub use is safe for you and your baby based on your individual circumstances.

GRADE: Acceptable

3.1.12. What Sleep Position Is Best during Pregnancy?

During pregnancy, it is recommended that women sleep on their left side as much as possible. This position allows for optimal blood flow to the uterus and kidneys, which can help to prevent swelling, back pain, and other discomforts associated with pregnancy.

Sleeping on the left side also helps to relieve pressure on the inferior vena cava, a large vein that runs along the right side of the body and can become compressed by the weight of the uterus when lying on the back. Compression of this vein can cause dizziness, shortness of breath, and a drop in blood pressure.

While sleeping on the left side is preferred, it is important to note that pregnant women should avoid sleeping on their backs after the first trimester. This is because as the uterus grows, it can put pressure on the vena cava and lead to the symptoms mentioned above.

Additionally, it is important for pregnant women to use pillows and other supports to help them stay comfortable while sleeping. This may include placing a pillow between the legs to alleviate pressure on the hips, or using a body pillow for additional support.

GRADE: Not Acceptable (Incomplete)

We found ChatGPT's sleep position response to be not acceptable because it is incomplete. Its omission of a prospective evaluation of maternal sleeping position published in 2019 has the potential to mislead patients, as supine or non-left sided sleep through 30 weeks was not associated with adverse pregnancy outcomes [18].

3.2. ChatGPT Reference Results

ChatGPT offered verified references for its responses to questions regarding caffeine, deli meat, laser, marijuana use, and sleep position. It provided no specific references for its responses on alcohol, hair dye, hot tub, vaping, and yoga. ChatGPT provided a ghost reference for its response on sushi and an irrelevant reference for its response on air travel. Overall, ChatGPT had references deficiencies for 58% (7/12) of the questions. When ChatGPT could not provide references, it would respond with a disclaimer: *I apologize for any confusion, but as an AI language model, I don't have direct access to external references or a browsing capability. My responses are generated based on a mixture of licensed*

data, data created by human trainers, and publicly available data. I have been trained on a wide variety of sources, including books, websites, and other texts, to develop a broad understanding of human language, but I do not have access to specific references for individual statements.

4. Discussion

ChatGPT and similar LLM technologies have the potential to link the world's knowledge to anyone with internet access. Since its release, this potential has been excitedly explored not only by academics, researchers, and clinicians but by the public as well. Initially, ChatGPT-3.5 demonstrated its ability to write e-mails, essays, poems, and songs. It tackled a variety of advanced level examinations including the GRE (Graduate Record Examination), SAT (Scholastic Aptitude Test), and LSAT (Law School Admission Test) and excelled. It passed both the Uniform Bar Exam and USMLE (United States Medical Licensing Examination) [10] [25]. This chatbot also satisfied a panel of experts in medicine and law on its ability to provide information about "virtually any topic in obstetrics and gynecology" [5]. These are astonishing achievements. Rapid progression of chatbots provides a hopeful view of a world where one can ask for and receive a near-human response from a machine.

Assessment of the accuracy of ChatGPT in the medical field, however, has produced mixed results. Yeo *et al.* evaluated ChatGPT on 164 hepatology questions, noting that 75% of responses were correct but less than half were comprehensive [12]. Drake (2023) demonstrated that ChatGPT was incorrect regarding the use of digital rectal exams in prostate cancer screening [26]. Alkaissi (2023) noted that ChatGPT described liver involvement in a rare glycogen storage disorder, an association that, in reality, has not yet been reported [27]. Multiple studies have also found that ChatGPT will "hallucinate" or make up references, in addition to providing article links and DOIs that do not exist [7].

User beware! LLMs can generate such coherent and "eloquent" responses so as to mislead the user. Human reviewers were asked to differentiate between ChatGPT-generated scientific abstracts and the original abstract and they struggled, incorrectly identifying more than 30% of AI-generated abstracts as original and 14% of original abstracts as AI-generated [28]. The issue of reference verification has been identified and studied as well.

We set out to verify that ChatGPT could be used as an accurate and reliable source of information for pregnant women without specific medical training or expertise. In regard to ChatGPT's responses to commonly asked pregnancy questions, 50% were not acceptable, being either incomplete or incorrect. In regard to references, more than 50% were deficient, with ChatGPT providing either broken, irrelevant, non-existent, or simply no references at all.

Pregnant women turn to the internet for information regarding their pregnancy. ChatGPT is an attractive and convenient tool that can compile internet resources in a digestible format. However, it is essential to emphasize that it can

be a flawed tool in the hands of the lay person without medical knowledge. Our analysis of ChatGPT regarding commonly asked pregnancy questions demonstrates the potential to mislead pregnant women given that its responses on air travel, laser, sushi, vaping, yoga, and sleep position were deemed to be not accurate. A chatbot response that differs from evidence-based information may cause confusion and stress, and it has the potential to cause harm.

New technologies may be powerful and exciting; however, they must be vetted prior to widespread use and integration into society. We concur with Grünebaum *et al.* that chatbots are a work in progress. Comparison of ChatGPT to the TI-30 calculator, the technological advancement of its time, reveals two very different stories. This minicalculator, launched almost 50 years ago, became the most popular scientific calculator for junior high and high school use in the United States with an estimated 15 million units being used between 1976 and 1983 [29]. Minicalculators were endorsed by the National Council of Teachers of Mathematics (NCTM) in 1975 and their use in the classroom was recommended [30]. Such endorsement gave the "end-user" confidence in the capability and accuracy of the technology. In contrast, ChatGPT carries no such endorsement. The amount of prerequisite knowledge or expertise necessary to operate a minicalculator pales in comparison to that required to effectively differentiate fact from fiction, plausible from non-plausible, and the safe from unsafe responses from ChatGPT.

The World Health Organization (WHO) recently issued an advisory statement for patients, healthcare providers, health systems, and policymakers cautioning that "precipitous adoption of untested systems could lead to errors by healthcare workers, cause harm to patients, erode trust in AI and thereby undermine (or delay) the potential long-term benefits and uses of such technologies around the world" [31]. Ironically, ChatGPT commented on this very topic when given the following prompt: **What are some of the concerns regarding use of ChatGPT in healthcare?**

*To address these concerns, healthcare providers and AI developers need to implement strict protocols for using AI models responsibly and transparently. Combining AI's capabilities with human expertise and oversight can help optimize patient care while mitigating potential risks. Regular updates and advancements in AI technology may also lead to more refined and accurate AI systems for healthcare applications (*https://www.chat.OpenAI.com*, accessed July 27, 2023).*

Our findings underscore the importance of proper testing and evaluation of ChatGPT and other LLMs prior to widespread adoption by healthcare workers and patients. Our qualitative analysis raises significant concerns regarding the use of ChatGPT as a source of information for commonly asked pregnancy questions. Given that ChatGPT's user experience has been described as "frictionless" with a "wow" factor along with a small learning curve, the best methodologies to evaluate LLMs must be explored [32]. Its potential role in medical decision making and as an independent source of information for those seeking

medical advice and reassurance warrants further vetting. At this stage, ChatGPT should be approached and used with caution.

The strengths of our study are the review panel and grading system. Our panel was a diverse team of medical professionals and students with sufficient knowledge, perspective, and interest in new technology to provide a fair but rigorous review of ChatGPT's performance. Three reviewers had a specific interest in AI and one reviewer actually used the TI-30 while in junior high school. Our three-tiered grading system was clearly defined and should prove easy to use in future studies.

Several limitations must be considered. We did not assess other LLMs, such as Google's Bard, that have recently become available to the public, nor did we assess the most current version of ChatGPT, Version 4. ChatGPT-3.5 was chosen because it was freely available at the time of this study and required no subscription fee. We acknowledge the performance improvements by newer versions of ChatGPT with "GPT-4" outperforming "GPT-3.5" on the SAT, scoring a 89th-percentile on the math section versus 70th-percentile, respectively [25]. In addition, we acknowledge that the responses analyzed for this review may not be the same responses as those in the future. We did not confirm reproducibility of responses within ChatGPT, although prior reports confirm strong reproducibility [12]. Despite our best intentions, evaluation and grading of ChatGPT's responses are subject to sampling bias as the commonly asked pregnancy questions we chose to query may not, in fact, be the "most common". Finally, the qualitative research approach, while appropriate, makes statistical analysis difficult. We look forward to the ideas and future work of others to confirm our findings and conclusions.

5. Conclusion

Our evaluation of ChatGPT's responses to commonly asked pregnancy questions confirms the concerns raised by others regarding both content and references. Since women currently use the internet as a source of information, it is likely that they will turn to AI to answer their pregnancy questions as well. While the debate continues regarding what AI is and how it is to be used, technological advances such as ChatGPT simply cannot be ignored. Subject expertise is required in order to evaluate ChatGPT output, and, therefore, we recommend that both professionals and laypersons approach this technology with caution.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper. The authors received no financial support for the research, authorship, and/ or publication of this article.

References

[1] Hu, K. (2023, February 2) ChatGPT Sets Record for Fastest-Growing User Base—

Analyst Note.
https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01

[2] American Medical Association (2023, March 3) ChatGPT Passed the USMLE. What Does It Mean for Med Ed?
https://www.ama-assn.org/practice-management/digital/chatgpt-passed-usmle-what-does-it-mean-med-ed

[3] Kung, T.H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., *et al*. (2023) Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models. *PLOS Digital Health*, **2**, e0000198.
https://doi.org/10.1371/journal.pdig.0000198

[4] Ayers, J.W., Poliak, A., Dredze, M., Leas, E.C., Zhu, Z., Kelley, J.B., Faix, D.J., Goodman, A.M., Longhurst, C.A., Hogarth, M. and Smith, D.M. (2023) Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. *JAMA Internal Medicine*, **183**, 589-596.
https://doi.org/10.1001/jamainternmed.2023.1838

[5] Grünebaum, A, Chervenak, J., Pollet, S.L., Katz, A. and Chervenak, F.A. (2023) The Exciting Potential for ChatGPT in Obstetrics and Gynecology. *American Journal of Obstetrics and Gynecology*, **228**, 696-705. https://doi.org/10.1016/j.ajog.2023.03.009

[6] Nolan, B. (2023) Two Professors Who Say They Caught Students Cheating on Essays with ChatGPT Explain Why AI Plagiarism Can Be Hard to Prove. Business Insider.
https://www.businessinsider.com/chatgpt-essays-college-cheating-professors-caught-students-ai-plagiarism-2023-1

[7] Sanchez-Ramos, L., Lin, L. and Romero, R. (2023) Beware of References When Using ChatGPT as a Source of Information to Write Scientific Articles. *American Journal of Obstetrics and Gynecology*, **229**, 356-357.
https://doi.org/10.1016/j.ajog.2023.04.004

[8] OpenAI. Introducing ChatGPT. 2022. https://openai.com/blog/chatgpt

[9] Declercq, E.R., Sakala, C., Corry, M.P., Applebaum, S. and Herrlich, A. (2014) Major Survey Findings of Listening to Mothers(SM) III: Pregnancy and Birth: Report of the Third National U.S. Survey of Women's Childbearing Experiences. *The Journal of Perinatal Education*, **23**, 9-16. https://doi.org/10.1891/1058-1243.23.1.9

[10] Popp, T. (2023, April 26) Alien Minds, Immaculate Bullshit, Outstanding Questions. *The Pennsylvania Gazette*.
https://thepenngazette.com/alien-minds-immaculate-bullshit-outstanding-questions/

[11] Kelly, S., Kaye, S.A. and Oviedo-Trespalacios, O. (2023) What Factors Contribute to the Acceptance of Artificial Intelligence? A Systematic Review. *Telematics and Informatics*, **77**, Article ID: 101925. https://doi.org/10.1016/j.tele.2022.101925

[12] Yeo, Y.H., Samaan, J.S., Ng, W.H., Ting, P.S., Trivedi, H., Vipani, A., *et al*. (2023) Assessing the Performance of ChatGPT in Answering Questions Regarding Cirrhosis and Hepatocellular Carcinoma. *Clinical and Molecular Hepatology*.
https://doi.org/10.1101/2023.02.06.23285449

[13] American College of Obstetricians and Gynecologists (2018) Air Travel during Pregnancy. ACOG Committee Opinion No. 746. *Obstetrics & Gynecology*, **132**, e64-e66.
https://doi.org/10.1097/AOG.0000000000002757

[14] Wilkerson, E.C., Van Acker, M.M., Bloom, B.S. and Goldberg, D.J. (2019) Utilization of Laser Therapy during Pregnancy: A Systematic Review of the Maternal and Fetal Effects Reported from 1960 to 2017. *Dermatologic Surgery*, **45**, 818-828.

https://doi.org/10.1097/DSS.0000000000001912

[15] ACOG Practice Advisory (2017) Update on Seafood Consumption in Pregnancy.

[16] American College of Obstetricians and Gynecologists (2020) Tobacco and Nicotine Cessation during Pregnancy. ACOG Committee Opinion No. 807. *Obstetrics & Gynecology*, **135**, e221-e229. https://doi.org/10.1097/AOG.0000000000003822

[17] Polis, R.L., Gussman, D. and Kuo, Y.H. (2015) Yoga in Pregnancy: An Examination of Maternal and Fetal Responses to 26 Yoga Postures. *Obstetrics & Gynecology*, **126**, 1237-1241. https://doi.org/10.1097/AOG.0000000000001137

[18] Silver, R.M., Hunter, S., Reddy, U.M., *et al.* (2019) Prospective Evaluation of Maternal Sleep Position through 30 Weeks of Gestation and Adverse Pregnancy Outcomes. *Obstetrics & Gynecology*, **134**, 667-676.
https://doi.org/10.1097/AOG.0000000000003458

[19] ACOG. Fetal Alcohol Spectrum Disorders FAQs.
https://www.acog.org/programs/fasd/fasd-faqs

[20] American College of Obstetricians and Gynecologists (2010) Moderate Caffeine Consumption during Pregnancy. Committee Opinion No. 462. *Obstetrics & Gynecology*, **116**, 467-468. https://doi.org/10.1097/AOG.0b013e3181eeb2a1

[21] ACOG. Listeria and Pregnancy.
https://www.acog.org/womens-health/faqs/listeria-and-pregnancy

[22] ACOG. Is It Safe to Dye My Hair during Pregnancy?
https://www.acog.org/womens-health/experts-and-stories/ask-acog/is-it-safe-to-dye-my-hair-during-pregnancy

[23] American College of Obstetricians and Gynecologists (2017) Marijuana Use during Pregnancy and Lactation. Committee Opinion No. 722. *Obstetrics & Gynecology*, **130**, e205-e209. https://doi.org/10.1097/AOG.0000000000002354

[24] American Academy of Pediatrics, American College of Obstetricians and Gynecologists (2017) Guidelines for Perinatal Care. 8th Edition.

[25] Varanasi, L. (2023) AI Models like ChatGPT and GPT-4 Are Acing Everything from the Bar Exam to AP Biology. Here's a List of Difficult Exams both AI Versions Have Passed. Business Insider.
https://www.businessinsider.com/list-here-are-the-exams-chatgpt-has-passed-so-far-2023-1

[26] Drake, L. (2023, May 2) We Fact-Checked ChatGPT's Medical Advice. Healthnews.
https://healthnews.com/news/we-fact-checked-chatgpt-medical-advice

[27] Alkaissi, H. and McFarlane, S.I. (2023) Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. *Cureus*, **15**, e35179.
https://doi.org/10.7759/cureus.35179

[28] Gao, C.A., Howard, F.M., Markov, N.S., Dyer, E.C., Ramesh, S., Luo, Y. and Pearson, A.T. (2023) Comparing Scientific Abstracts Generated by ChatGPT to Real Abstracts with Detectors and Blinded Human Reviewers. *NPJ Digital Medicine*, **6**, Article No. 75. https://doi.org/10.1038/s41746-023-00819-6

[29] TI-30. In: Wikipedia. 2023.
https://en.wikipedia.org/w/index.php?title=TI-30&oldid=1141113048

[30] Texas Instruments Incorporated (1977) TI-30 Student Math Kit.

[31] World Health Organization (2023, May 23) WHO Calls for Safe and Ethical AI for Health.
https://www.who.int/news/item/16-05-2023-who-calls-for-safe-and-ethical-ai-for-health

[32] Blagic, D. (2022, December 20) Why Is the User Experience of Chatgpt So Powerful? Medium.
https://uxdesign.cc/why-is-the-user-experience-of-chatgpt-so-powerful-509e803e0122

[33] Piloto, C. (2022, December 26) Artificial Intelligence vs Machine Learning: What's the Difference? MIT Professional Education.
https://professionalprograms.mit.edu/blog/technology/machine-learning-vs-artificial-intelligence

[34] What Is a Chatbot? IBM. https://www.ibm.com/topics/chatbots

[35] Ji, Z., Lee, N., Frieske, R., *et al.* (2023) Survey of Hallucination in Natural Language Generation. *ACM Computing Surveys*, **55**, Article No. 248.
https://doi.org/10.1145/3571730

[36] (2023) Prepare for Truly Useful Large Language Models. *Nature Biomedical Engineering*, **7**, 85-86. https://doi.org/10.1038/s41551-023-01012-6

[37] What Is Natural Language Processing? IBM.
https://www.ibm.com/topics/natural-language-processing

[38] About. OpenAI. https://openai.com/about

Appendix

Glossary of Terminology Relating to Artificial Intelligence

Artificial intelligence (AI)	The ability for computers to imitate cognitive human functions such as learning and problem-solving. Utilized in chatbots, cell phones, vehicles, banking, and more, an AI computer system uses math and logic to simulate reasoning that people use to learn from new information and make decisions [33].
Chatbot	A computer program utilizing artificial intelligence to understand the user's question and automate responses to them, simulating human conversation. A familiar example includes shopping chatbots that answer customer service and product information questions on the company's website. More recently, companies like OpenAI, Google, and Microsoft have been developing more advanced chatbots with large language modeling technology [34].
Digital Assistant	A computer program using artificial intelligence and enhanced with natural language processing to interact with the user and assist with daily activities (e.g. setting alarms, making calls, getting directions). Examples include Apple's Siri, Google Assistant, and Amazon Alexa. Alternative terms include "virtual assistant" and "virtual agent" [34].
Hallucination	In the context of chatbots, a hallucination is the generation of "content that is nonsensical or unfaithful to the provided source content." In other words, the program "makes up" information. This error may be due to discrepancies in the data used for program training or the training itself [35].
Large language model (LLM)	A type of AI system, trained on a vast amount of data and parameters, seeking to model human language. LLMs have learned "structural, relational, and semantic language patterns" that allow them to predict the best next word in a piece of text [36]. ChatGPT is an example of an LLM.
Machine learning	The application of teaching computers to extract patterns from collected data for new tasks that they may not have completed before. A real-world example is the ability of music streaming platforms to recommend custom songs and artists based on your past listening history, using data from what other users with similar taste have listened to [33].
Natural language processing (NLP)	Branch of artificial intelligence aimed to program computers to fully "understand" human language to accurately complete the user's intent and sentiment [37].
OpenAI	An American artificial intelligence company, founded in 2015, that created ChatGPT. ChatGPT is version 3.5 of OpenAI's chatbot, and was made freely available to the public in November 2022. The most recent version of their chatbot is known as GPT-4, released in March 2023 [38].
Query	A request for data or information. In the context of AI and chatbots, the query is the user's input or question that is entered into the computer system.