Evaluating the Accuracy, Completeness, and Readability of Artificial Intelligence-Generated Information on Commonly Promoted Fertility Supplements in English and Spanish

Lucia Eade, B.S., Andrea Parra, M.D., Samantha Alvarado, M.D., Ashlyn Parkhurst, M.D., Nicholas Stansbury, M.D.
The University of Texas Health Science Center at San Antonio

Introduction

Patients increasingly turn to artificial intelligence (AI) for guidance on fertility supplements. While these tools may offer rapid, accessible information, their accuracy, completeness, and safety counseling have not been well characterized—particularly across languages. Given that Spanish-speaking patients already face disparities in reproductive health access, AI-generated responses may alter this gap depending on their quality. This study objective is to evaluate ChatGPT's English- and Spanish-language responses on 14 commonly used fertility supplements to assess accuracy, completeness, readability, and safety counseling.

Materials and Methods

Fourteen fertility supplements commonly promoted across consumer-facing resources, including online marketplaces and fertility-focused websites, were selected based on recurring recommendations. For each supplement, five standardized prompt categories were developed to query: (1) intended use, (2) efficacy, (3) risks, (4) pregnancy safety, and (5) patient counseling considerations. Each prompt was submitted in both English and Spanish to ChatGPT (version 5.5), yielding 140 total queries. All queries were submitted as independent, single-turn interactions to avoid carryover effects from conversational context.

A reference standard was compiled by the study team from multiple evidence-based sources to serve as the benchmark against which ChatGPT responses were evaluated for factual accuracy and completeness.

Three bilingual reviewers independently evaluated each ChatGPT response. Reviewers were blinded to the language order in which responses were generated to limit bias in scoring. Each response was rated using a 5-point Likert scale across four domains. Inter-rater agreement among the three reviewers was assessed using agreement within one Likert point (±1) for each domain.

Domain	English Mean Score	Spanish Mean Score	Clinical Interpretation
Accuracy	3.79	3.64	English responses were more factually aligned with reference standard
Completeness	3.43	3.21	Spanish responses had more omitted counseling elements
Safety counseling	Not separately calculated	3.31	Safety counseling was limited and lowest in Spanish
Readability	3.93	3.58	Responses were generally readable, with higher English scores

Table 1. Overall domain scores by language

Finding	Supplement(s)	Interpretation
Highest overall performance	Folic acid, myo-inositol	More consistent evidence-based counselling in both languages
Lowest completeness score	D-aspartic acid (Spanish-language)	Completeness score 2.67
Most safety-sensitive supplement categories	DHEA, iodine, ashwagandha, melatonin, zinc, D-aspartic acid	Greater need for contraindication, pregnancy, endocrine, and medication-interaction counselling
Supplements prone to benefit overstatement	CoQ10, DHEA, melatonin, maca, D-aspartic acid	Responses risked conflating biologic plausibility with proven fertility benefit

Table 2. Supplements with notable scoring patterns

Results

A total of 140 AI-generated responses were evaluated. English responses scored higher than Spanish responses across all evaluated domains, including accuracy, completeness, and readability. Safety counseling remained limited in both languages and was lowest in Spanish. Inter-rater reliability showed close reviewer agreement, with 100% agreement within one point across all domains.

Conclusion

AI-generated fertility supplement information is generally readable but shows meaningful gaps in completeness and safety counseling, particularly in Spanish. These findings suggest that AI tools may reinforce language-based disparities in reproductive health information access. Clinicians should counsel patients that AI-generated content does not replace individualized medical guidance, particularly when supplements may interact with fertility treatments, pregnancy, endocrine disorders, or hormone-sensitive conditions.

References

1. Lundin K, Bentzen JG, Bozdag G, Ebner T, Harper J, Le Clef N, et al. Good practice recommendations on add-ons in reproductive medicine. Hum Reprod. 2023;38(11):2062-2104.
2. Kamath MS, Mascarenhas M, Franik S, Liu E, Sunkara SK. Clinical adjuncts in in vitro fertilization: a growing list. Fertil Steril. 2019;112(6):978-986.
3. Nardo LG, El-Toukhy T, Stewart J, Balen AH, Potdar N. Adjuvants in IVF—evidence for what works and what does not work. Ups J Med Sci. 2020;125(2):144-151.
4. Shirlow R, Healey M, Volovsky M, MacLachlan V, Vollenhoven B. The effects of adjuvant therapies on embryo transfer success. J Reprod Infertil. 2017;18(4):368-378.