

## Prompt engineering for health equity: A modeled AI intervention to improve online Spanish-language orthopedic patient resources

Carrie Reaver, Elisa Victoria Carrillo, Daniel E Pereira, Carolena Rojas Marcos, Juan M Colazo, Charles A Goldfarb

### INTRODUCTION:

**Background:** Spanish-speaking patients in the U.S. face persistent barriers to health information due to the limited availability and poor readability of Spanish-language educational resources on academic orthopedic websites. This disparity is further compounded by a low representation of Spanish-speaking orthopedic surgeons. Generative artificial intelligence (AI) presents a promising solution to address this gap. This study modeled a practical, scalable AI intervention to generate accessible Spanish-language orthopedic educational materials that institutions can readily adopt to address language disparities.

### METHODS:

**Methods:** We conducted a cross-sectional analysis of the 203 U.S. academic orthopedic institutions listed on the 2024 ERAS directory to assess the availability of Spanish-language patient resources on each institution's website. We then modeled an AI intervention using 30 articles from Washington University's orthopedic patient library using ChatGPT-4o, as currently no Spanish resources are available. Original English-language articles were translated to Spanish and rewritten to meet the NIH/AMA-recommended sixth-grade reading level. We approximated a direct translation of English articles with Google Translate to serve as our "original" comparison against AI-generated Spanish materials. Prompt development was performed using proven strategies including context/persona building, few-shot prompting, structured chain prompting, and iterative testing until satisfactory outputs were achieved (Fig. 1). Readability was assessed using validated readability tests in both English and Spanish. A modified DISCERN tool was used to assess content quality and translation quality by two native Spanish-speaking orthopedic surgeon graders. Readability and quality scores before and after ChatGPT intervention were compared with paired t-tests. Correlation between changes in reading grade level and quality was assessed with Pearson's correlation. Significance was set to  $p < 0.05$  for all statistics.

### RESULTS:

**Results:** Of the 203 institutions included in our search, 142 (70%) provided patient educational resources in English on their websites. Of these 142 institutions, only 17 (12%) provided Spanish translations of the material. The cumulative average reading grade level of WashU's patient articles was  $10.7 \pm 2.3$  in English. Google Translate was used on the original articles with a cumulative average reading grade level of  $10.1 \pm 1.7$  in Spanish. ChatGPT significantly reduced the reading level in both languages, achieving an average of  $6.9 \pm 1.5$  in English ( $p < 0.01$ ) and  $6.8 \pm 1.2$  in Spanish ( $p < 0.01$ ) (Fig. 2). ChatGPT-generated Spanish articles demonstrated either improvement (grader 1,  $p < 0.05$ ) or noninferiority (grader 2,  $p = 0.37$ ) in DISCERN content quality scores following ChatGPT intervention. The quality of translation was scored significantly higher by both graders for ChatGPT content (both graders  $p < 0.05$ ). No significant negative correlation was observed between improvements in readability and content quality scores for either grader (Grader 1:  $r = 0.34$ ,  $p = 0.23$ ; Grader 2:  $r = 0.12$ ,  $p = 0.67$ ).

### DISCUSSION AND CONCLUSION:

**Conclusions:** Our modeled intervention demonstrates the effectiveness of generative AI for enhancing the accessibility and quality of online Spanish-language orthopedic education materials. We also introduce prompt engineering strategies that support the creation of high-quality outputs and build competency in AI-driven patient care. Academic orthopedic departments may consider leveraging these strategies to address language barriers and promote more equitable, accessible patient education.

