Regulating Linguistic Factors to Optimize AI Simplification of Orthopedic Text

Saman Andalib[1], Sean Solomon[1], Bryce Picton[1], Aidin Spina[1], John Scolaro[1], Ariana Nelson[1]
[1]University of California, Irvine, School of Medicine, CA
sandalib@hs.uci.edu

Disclosures: SA, SS, BP, AS, JS, and AN have no disclosures.

INTRODUCTION: Patient education materials (PEMs) in orthopedics frequently utilize complex language, limiting their accessibility and comprehension by patients. Large language models (LLMs), a type of artificial intelligence, present a potential solution by automatically simplifying the text of PEMs. However, the factors that predict whether an LLM will successfully simplify a given PEM remain unknown. Elucidating these factors would allow providers to optimize PEMs for simplified output.

METHODS: 48 orthopedic PEMs were compiled and simplified using four LLMs: GPT-4, GPT-3.5, Claude 2, and LLaMa 2. The readability of each PEM was quantified before and after simplification using standard formulas for Flesch-Kincaid Reading Ease (FKRE) and Grade Level (FKGL). Correlations between inherent PEM text factors and output readability were analyzed. Feature importance analysis identified the most predictive simplification factors.

RESULTS SECTION: Moderate correlations existed between original text factors and output FKRE/FKGL. For output FKRE, text length correlated negatively with GPT-4 and GPT-3.5 (coefficients -0.23 to -0.38) but positively for Claude 2 (coefficients +0.28 to +0.29). For output FKGL, text length correlated positively for GPT-3.5 (coefficients +0.21 to +0.25) but negatively for Claude 2 (coefficients -0.29 to -0.31). Original FKRE was the most predictive factor for output FKRE (importance scores 0.28-0.45). Original FKGL and syllables per sentence were most predictive for output FKGL (importance scores 0.21-0.29).

DISCUSSION: The results suggest LLMs simplify PEMs differently based on original text qualities. Further research should clarify how providers can tailor PEMs to optimize simplification success for a given LLM. For instance, materials with higher baseline readability may see greater improvement with GPT-4 and GPT-3.5, while Claude 2 may work better for more complex PEMs. Findings provide early insights into how factors like baseline readability, word complexity, and syntax may be regulated to improve LLM simplification. However, additional work is needed to determine the ideal protocol for optimizing PEMs for simplified output depending on the chosen LLM.

SIGNIFICANCE/CLINICAL RELEVANCE: Increasing PEM accessibility through AI-driven simplification has immense potential to improve patient education. These findings represent an important initial step toward determining how medical texts can be tailored to enhance comprehension.

Correlation of Original Text Factors with FKRE Scores

Correlation of Original Text Factors with FKGL Scores