

Supporting Information

Classifying Free Texts into Pre-Defined Sections Using AI in Regulatory Documents: A Case Study with Drug Labeling Documents

Magnus Gray¹, Joshua Xu¹, Weida Tong¹, Leihong Wu^{1*}

¹Division of Bioinformatics & Biostatistics, National Center for Toxicological Research, US Food and Drug Administration, 3900 NCTR Rd., Jefferson, AR 72079 USA

*Corresponding Author:

Leihong Wu, Division of Bioinformatics & Biostatistics, National Center for Toxicological Research, US Food and Drug Administration, 3900 NCTR Rd., Jefferson, AR 72079 USA

Email: Leihong.Wu@fda.hhs.gov

Table of Contents

Non-PLR & SmPC Shapley Additive Explanation Analysis	S2
Figure S1: Non-PLR Format Shapley Additive Explanations	S2
Figure S2: SmPC Format Shapley Additive Explanations	S3

Non-PLR & SmPC Shapley Additive Explanation Analysis

The two collections of charts below (*Figure S1* and *Figure S2*) document words which were the most influential in the four major categories in Non-PLR and SmPC drug labeling formats, respectively. While many of the selected tokens varied greatly between the three formats of drug labeling documents (i.e., PLR, Non-PLR, and SmPC), several patterns were noted. For instance, regarding the “Indications & Usage” section, each format was heavily influenced by the word “Indicated;” and in the “Warnings & Precautions” section, each classification was largely influenced by the term “Monitor.” Thus, this Shapley Additive Explanation analysis showed that, while the three formats were fundamentally different, many of the same or similar-meaning words influenced the models’ decisions when predicting the section endpoint for a given drug labeling sentence.

Figure S1: Non-PLR Format Shapley Additive Explanations

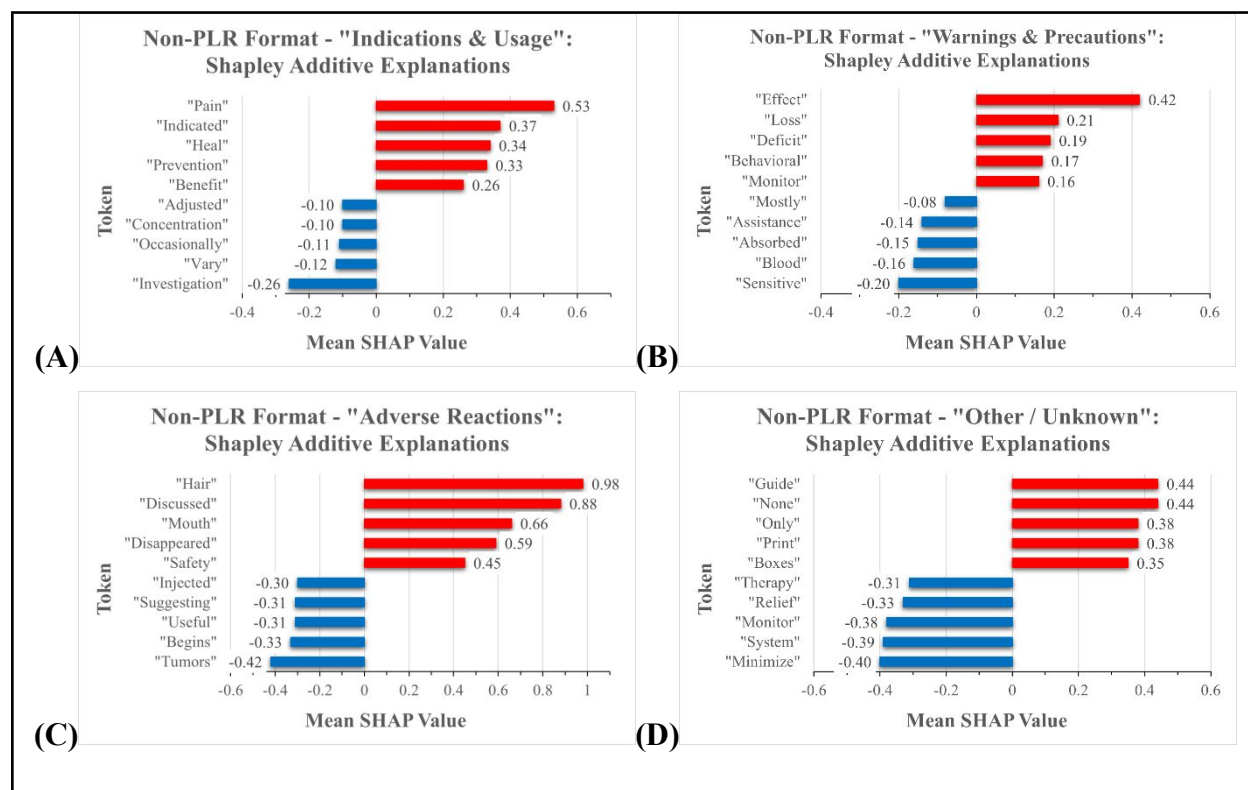


Figure S1: Non-PLR Format Shapley Additive Explanations. (A) Indications & Usage. (B) Warnings & Precautions. (C) Adverse Reactions. (D) Other/Unknown.

Figure S2: SmPC Format Shapley Additive Explanations

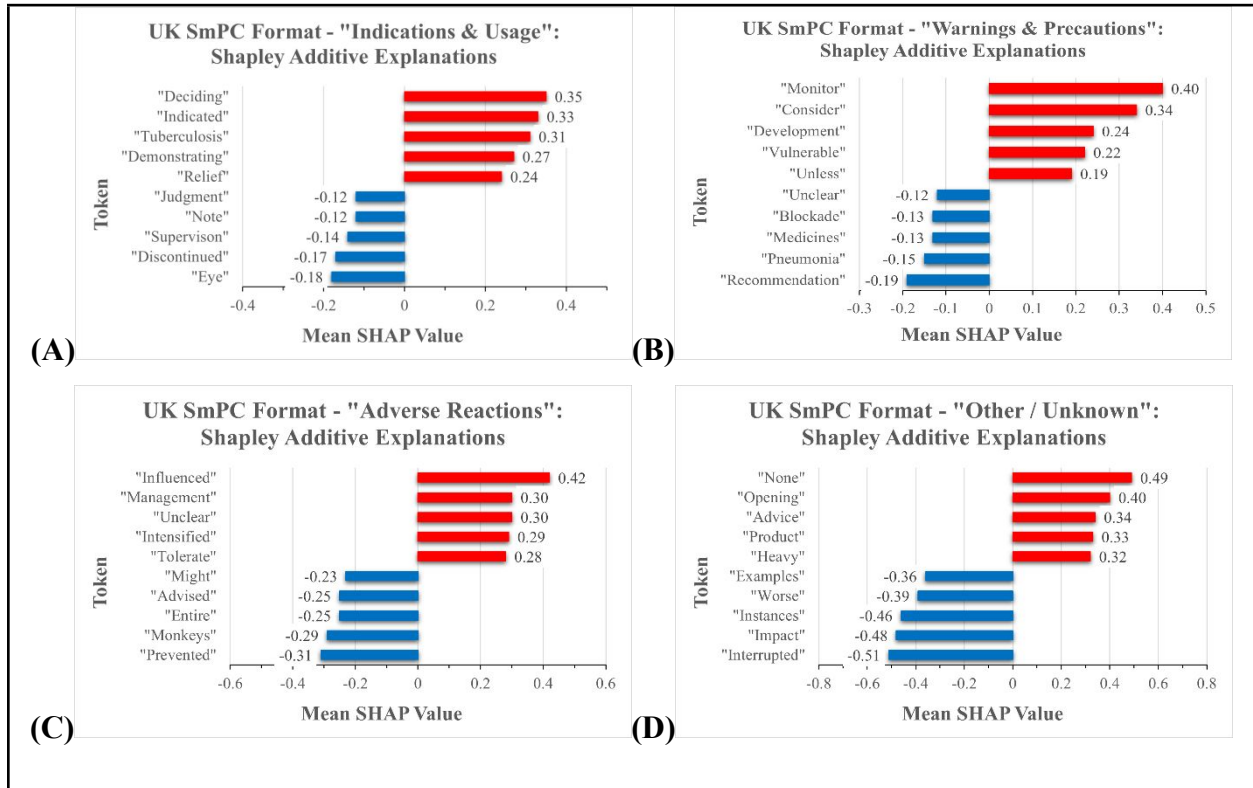


Figure S2: SmPC Format Shapley Additive Explanations. (A) Indications & Usage. (B) Warnings & Precautions. (C) Adverse Reactions. (D) Other/Unknown.