## **Supporting Information**

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

### Magnus Gray<sup>1</sup>, Joshua Xu<sup>1</sup>, Weida Tong<sup>1</sup>, Leihong Wu<sup>1\*</sup>

<sup>1</sup>Division 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
	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.

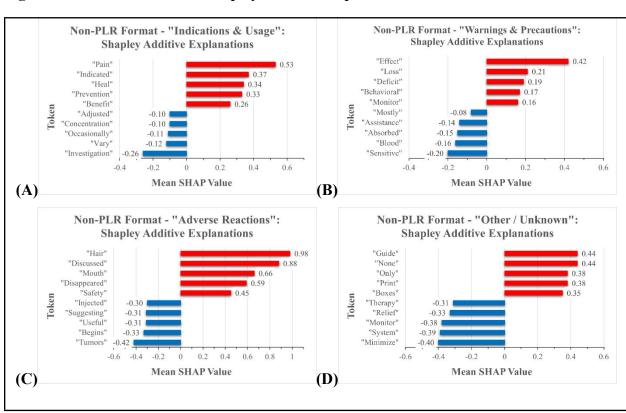


Figure S1: Non-PLR Format Shapley Additive Explanations

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

UK SmPC Format - "Warnings & Precautions": UK SmPC Format - "Indications & Usage": **Shapley Additive Explanations Shapley Additive Explanations** "Deciding" "Consider" "Development" "Tuberculosis" "Vulnerable" "Demonstrating" "Relief" "Unclear' "Judgment" "Blockade" -0.13 "Note" "Supervison" "Medicines' -0.13 "Discontinued" "Pneumonia' -0.15 "Eye" "Recommendation" 0.4 -0.3 -0.2 -0.1 0.1 Mean SHAP Value Mean SHAP Value **(A) (B)** UK SmPC Format - "Adverse Reactions": UK SmPC Format - "Other / Unknown": **Shapley Additive Explanations Shapley Additive Explanations** "None" "Influenced" "Management" 0.30 "Opening" "Unclear" 0.30 "Advice" "Intensified" 0.29 "Product" 0.33 "Tolerate" "Might" 0.28 "Heavy" "Examples" -0.36 -0.23"Advised" "Worse" -0.39 -0.25 "Entire" -0.25 "Instances" -0.46 "Monkeys" -0.29 "Impact" -0.48 "Prevented" "Interrupted" -0.31 -0.51 0.2 -0.6 0.2 -0.6 -0.40.4 0.6 -0.4 -0.2 0.4 0.6 Mean SHAP Value Mean SHAP Value **(C) (D)** 

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.